

INTERNATIONAL JOURNAL OF MEDICAL BIOCHEMISTRY

YEAR 2026 VOLUME 9 ISSUE 4

ORIGINALS

Low AMH, inflammation, and nutrition

Dundar Akin, et al.

Correlates of low fecal elastase-1 levels

Hacihasanoglu Cakmak, et al.

Comparison of ACTH results in K2 and K3 EDTA tubes

Ustundag, et al.

Verification of Maccura F810 hematology analyzer

Cinpolat, et al.

HBT results and analytical change limits in thyroid testing

Turkal, et al.

Machine learning for abnormal liver biochemistry

Demir, et al.

Predictors of emergency department specimen rejection

Goren, et al.



EDITOR-IN-CHIEF **Prof. DILDAR KONUKOGLU, MD.**
Istanbul University-Cerrahpasa, Cerrahpasa Faculty of Medicine, Istanbul, Türkiye

ASSOCIATE EDITOR **Prof. NECIP ILHAN, MD.**
Firat University, Faculty of Medicine, Elazığ, Türkiye

EDITORIAL ASSISTANT **E. CUNEYT CANBULAT, MD.**
General Manager of KBUDEK External Quality Control Programme, Istanbul, Türkiye

PUBLISHING MANAGER **Prof. DILDAR KONUKOGLU, MD.**
Istanbul University-Cerrahpasa, Cerrahpasa Faculty of Medicine, Istanbul, Türkiye

PUBLISHING MANAGER ASSISTANT **FIKRET ERSAN TALU, MD.**
Secretary of Association of Clinical Biochemistry Specialists, Türkiye

LINGUISTIC EDITOR **Prof. UZAY GORMUS DÉGRIGO, MD., PhD.**
Karolinska Institutet, Solna, Sweden

EXTERNAL EDITORS **ERDINC SEZGIN, PhD.**
Karolinska Institutet, Department of Women's and Children's Health, Scilifelab, Sweden

Prof. EVIN ADEMOGLU, MD.
Istanbul University, Faculty of Medicine, Department of Medical Biochemistry,
Istanbul, Türkiye

ADVISORY BOARD

Prof. KHOSROW ADELI, MD., PhD.
Toronto University, The Hospital for Sick Children and
Departments of Laboratory Medicine and Pathobiology,
Biochemistry, and Physiology, Ontario, Canada

Prof. ANGELO AZZI, MD., PhD.
Tufts University, Jean Mayer USDA Human Nutrition
Research Ctr. on Aging (HNRCA), Boston, Massachusetts

Prof. BANU ISBİLEN BASOK, MD.
University of Health Sciences, İzmir Faculty of Medicine, Dr.
Behcet Uz Child Disease and Pediatric Surgery Training and
Research Hospital, İzmir, Türkiye

Prof. ETIENNE CAVALIER, MD.
Centre Hospitalier Universitaire de Liège, Specialist in Laboratory
Medicine, EuSplM, University of Liège, Liège, Belgium

Prof. AMITAVA DASGUPTA, PhD.
University of Texas Health Sciences Center, Department of
Pathology and Laboratory Medicine, Houston, USA

Prof. KAYA EMERK, PhD.
Istanbul Kent University, Faculty of Dental Medicine,
Department of Biochemistry, Istanbul, Türkiye

Prof. AMMAD AHMAD FAROOQI, MD.
Institute of Biomedical and Genetic Engineering (IBGE),
Islamabad, Pakistan

Prof. KYM FRANCIS FAULL, PhD.
University of California Los Angeles, Pasarow Mass
Spectrometry Laboratory, Los Angeles, USA

Prof. ASUMAN GEDIKBASI, MD., PhD.
Istanbul University, Institute of Child Health, Pediatric Basic
Sciences, Medical Genetics, Istanbul, Türkiye

Prof. UZAY GORMUS DÉGRIGO, MD., PhD.
Karolinska Institutet, Department of Medical Biochemistry
and Biophysics, Solna, Sweden

Prof. ASLIHAN GURBUZ BOLKAN, MD.
Ankara University, Faculty of Medicine, Department of
Biochemistry, Ankara, Türkiye

Prof. NEVIN ILHAN, MD., PhD.
Firat University, Faculty of Medicine, Department of Medical
Biochemistry, Elazığ, Türkiye

Prof. FERRUH K. ISMAN, MD.
Istanbul Medeniyet University, Goztepe Training and
Research Hospital, Department of Clinical Chemistry,
Istanbul, Türkiye

Prof. CIGDEM KARAKUKCU, MD.
Erciyes University, Faculty of Medicine, Department of
Biochemistry, Kayseri, Türkiye

Prof. HUSEYİN KAYADIBI, MD.
Eskisehir Osmangazi University, Faculty of Medicine,
Department of Biochemistry, Eskişehir, Türkiye

Prof. GABOR L. KOVACS, MD., PhD, DSc.
University of Pécs, Faculty of Medicine, Department of
Laboratory Medicine, Pécs, Hungary

Prof. MINE KUCUR, MD.
Istanbul University Cerrahpasa, Cerrahpasa Faculty of
Medicine, Department of Biochemistry, Istanbul, Türkiye

Prof. ASİM OREM, MD.
Karadeniz Technical University, Faculty of Medicine, Trabzon,
Türkiye

Assoc. Prof. İSHAK OZEL TEKİN, MD. PhD.
Bulent Ecevit University, Faculty of Medicine, Department of
Immunology, Zonguldak, Türkiye

Prof. MAURO PANTEGHINI, MD.
Milan University, Faculty of Medicine, Department of Clinical
Biochemistry and Clinical Molecular Biology, Milano, Italy

Prof. JORGE L. SEPULVEDA, MD, PhD.
Columbia University, Department of Pathology and Cell Biology,
New York, USA

Prof. MUHİTTİN ABDULKADİR SERDAR, MD.
Acibadem Mehmet Ali Aydınlar University, Faculty of Medicine,
Department of Medical Biochemistry, Istanbul, Türkiye

Prof. ANA-MARIA SIMUNDIC, PhD.
Sveti Duh University, Department of Medical Laboratory
Diagnostics, University Hospital "Sveti Duh", Zagreb, Croatia

Prof. STEVEN SOLDIN, MD.
Deputy Director of Chemistry, Senior Scientist and Director
Post-Doctoral Training Program, Department of Laboratory
Medicine, NIH Clinical Center Bethesda, USA and Adjunct
Professor, Division of Endocrinology and Metabolism
Georgetown University, Department of Medicine, USA

Prof. FATMA TANELİ, MD.
Celal Bayar University, Faculty of Medicine, Department of
Biochemistry, Manisa, Türkiye

DOUGLAS THOMPSON, PhD.
Leeds University, Leeds Teaching Hospitals, Department of
Blood Science, Leeds, UK

Prof. TOMRIS OZBEN TOMASI, MD., PhD.
Akdeniz University, Faculty of Medicine, Department of
Biochemistry, Antalya, Türkiye

Prof. TURAN TURHAN, MD.
Ankara Bilkent City Hospital Health Application and
Research Center, Department of Medical Biochemistry,
Ankara, Türkiye

ALEXANDER ZOUGMAN, MD, PhD.
St. James's University Hospital, Clinical and Biomedical
Proteomics Group, Cancer Research UK Centre, Leeds, UK

VOLUME IX ISSUE 4 YEAR 2026

The Owner and Publishing Manager on behalf of the Association of
Clinical Biochemistry Specialists (Klinik Biyokimya Uzmanları Derneği)

Address: Maslak Mah. AOS 55. Sok. No: 2, 42 Maslak A
Blok Daire: 231 Sarıyer, İstanbul-Türkiye
Phone: +90 212 241 26 53
Fax: +90 212 241 26 54
e-mail: www.kbud.org.tr
web: info@kbud.org.tr

International Journal of Medical Biochemistry is a peer-reviewed journal published triannually.
Materials published in the Journal is covered by copyright 2025. All rights reserved.
This publication is printed on paper that meets the international standard ISO 9706:1994.
National Library of Medicine (USA) recommends the use of permanent, acid-free paper in the production of biomedical literature.



Publisher: KARE MEDIA
Address: Göztepe Mah. Fahrettin Kerim Gökay Cad.
No: 200 D: 2, Göztepe, Kadıköy, İstanbul
Phone: +90 216 550 61 11
Fax: +90 216 550 61 12
e-mail: info@karepb.com
web: www.kareyayincilik.com

Publication Type: International Periodical
Publication Date: October 2026
Printing: Filmevi Grafik Renk Ayrımı
Sistemleri San. Tic. LTD. ŞTİ.
İstanbul-Türkiye



ABOUT JOURNAL

AIM AND SCOPE

International Journal of Medical Biochemistry (IJMB) is a double-blind peer-reviewed, open-access journal of the Association of Clinical Biochemistry Specialists Türkiye.

International Journal of Medical Biochemistry (IJMB) publishes articles relating to clinical and experimental chemistry, molecular biology, genetics, therapeutic drug monitoring, toxicology, immunology, hematology and laboratory medicine with the focus on analytical and clinical investigation of laboratory tests used for diagnosis, prognosis, and monitoring of diseases.

BASIC PUBLICATION RULES

Authors are responsible for the accuracy of data. The Editorial Board of the International Journal of Medical Biochemistry and the Publisher adheres to the principles of the International Council of Medical Journal Editors (ICMJE), the World Association of Medical Editors (WAME), the Council of Science Editors (CSE), the Committee on Publication Ethics (COPE), the US National Library of Medicine (NLM), the World Medical Association (WMA) and the European Association of Science Editors (EASE).

DISCLAIMER

Statements or opinions expressed in the manuscripts published in IJMB reflect the views of the author(s) and not the opinions of the editors, the editorial board and the publisher; the editors, the editorial board and the publisher disclaim any responsibility or liability for such materials.

JOURNAL FREQUENCY

The International Journal of Medical Biochemistry published three issues per year from its establishment until 2024. The journal has started publishing four issues per year since 2025. The publication months are January, April, July, and October.

ABSTRACTING AND INDEXING

International Journal of Medical Biochemistry is indexed in TUBITAK TR Index (2019), CNKI (2019), TurkMedline (2019), Open Ukrainian Citation Index (2019), ProQuest (2019), CABI (2021), CEEAS (2021), EBSCO (2022), CAS (American Chemical Society) (2022), Directory of Open Access Journals - DOAJ (2022), Scopus (2023), Gale Cengage (2023), Sherpa Romeo (2024), ASCI (2024) and Web of Science - ESCI (2025), EZB (2025), OpenAlex (2025), Sudoc (2025), ZDB (2025).

ABBREVIATION

Int J Med Biochem

PUBLICATION FEE

Submission Fee

As of January 1, 2026, authors are required to pay a submission fee of USD 120 (including VAT) at the time of manuscript submission to the International Journal of Medical Biochemistry (IJMB). The submission process will not be initiated for manuscripts for which the submission fee has not been paid. This fee is charged to contribute to the costs of editorial handling. The submission fee is applied during the manuscript submission process and is non-refundable, regardless of the final editorial decision (acceptance or rejection). Kindly note that any bank transaction fees must be covered by the author.

Article Processing Charge (APC)

As of January 1, 2026, manuscripts submitted to the International Journal of Medical Biochemistry (IJMB) that are accepted for publication will be subject to an Article Processing Charge (APC) of USD 120 (including VAT). The publication process will not proceed unless the APC has been fully paid. This fee is collected to support the production, and publication of the accepted manuscripts. Kindly note that any bank transaction fees must be covered by the author.

Additional Information

Authors based in Türkiye may pay the above fees in Turkish Lira (TRY), calculated based on the USD–TRY exchange rate applicable at the time of payment. For any questions regarding payments, please contact us at kare@karepb.com.

ACCESS TO JOURNAL CONTENT

The abstracts and full texts of published articles can be accessed free of charge at www.internationalbiochemistry.com.



INSTRUCTIONS FOR AUTHORS

AIMS & SCOPE

International Journal of Medical Biochemistry publishes articles relating to clinical chemistry, molecular biology and genetics, therapeutic drug monitoring and toxicology, hematology, immunology and laboratory medicine with the focus on analytical and clinical investigation of laboratory tests used for diagnosis, prognosis, treatment and therapy, and monitoring of disease. The official language of the Journal is English. The journal will be published 4 times a year in print and electronically.

Main Topics

- Clinical Biochemistry
- Molecular Biology
- Clinical Haematology
- Clinical Immunology
- Drug Monitoring and Analysis
- Diagnostic Biomarkers
- Disease-Oriented Topics (Cardiovascular Disease, Cancers, Diabetes, Obesity, Genetic Disorders, Neurodegenerative Disease e.g.)
- Pediatric Biochemistry
- Inherited Metabolic Disorders
- Newborn Screening: Congenital and Genetic Disorders
- New Reagents, Instrumentation, Technologies and Methodologies

- Laboratory Medicine; Quality, Safety, translational laboratory
- Clinical Metrology

INSTRUCTIONS FOR AUTHORS

International Journal of Medical Biochemistry is published in accordance with the principles of independent, unbiased, and double-blind peer review. Research articles, review articles, short communications, case-reports, opinion papers, letter to editor, technical notes, editorials and article-commentaries that have not been published elsewhere, are published in the journal. **Four issues are released every year in January, April, July, and October.**

The journal evaluates only the manuscripts submitted through its online submission system on the web site <http://www.internationalbiochemistry.com>. Manuscripts sent by other means will not be accepted.

The primary conditions for the acceptance of manuscripts for publication are originality, scientific value and citation potential.

The manuscripts should be prepared in accordance with ICMJE Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (updated in May 2022 - <https://www.icmje.org/recommendations/>).

Authors are required to prepare manuscripts in accordance with the international guidelines* below.

Randomized Controlled Trial **	CONSORT (Consolidated Standards of Reporting Trials)
Non-Randomized Trial **	TREND (Transparent Reporting of Evaluations with Non-randomised Designs)
Trial Protocol	SPIRIT (Standard Protocol Items Recommendations for Interventional Trials)
Observational Epidemiologic Studies (cohort, case-control cross-sectional)	STROBE (the STrengthening the Reporting of OBservational studies in Epidemiology)
Systematic Reviews and Meta-Analyses	PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses)
Systematic Reviews and Meta-Analyses Protocol	PRISMA-P (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Protocol)
Qualitative Research	-SRQR (the Standards for Reporting Qualitative Research) -COREQ (Consolidated criteria for REporting Qualitative research: interviews and focus groups)
Methodological Studies (Translating and Adapting Tests)	ITC (International Test Commission) Guidelines for Translating and Adapting Tests
Methodological Studies (Developing Tests)	COSMIN (Consensus based Standards for the selection of Health Measurement Instruments- Study Design for Patient-reported outcome measurement instruments)
Case Report	CARE (CAse REport)

* Enhancing the QUALity and Transparency of Health Research (equator network).

** The International Journal of Medical Biochemistry encourages the registration of all clinical trials (randomized and non-randomized) via ClinicalTrials.gov or one of the registries of the WHO's International Clinical Trials Registry Platform (ICTRP). The name of the trial registry and the registration number together should be provided at the end of the abstract.

PEER REVIEW PROCESS

International Journal of Medical Biochemistry utilizes a double-blind peer review process where both the identities of reviewers and authors are anonymous. The Editor-in-Chief is responsible for the editorial and scientific content of the journal and makes the final decision.

Submitted manuscripts that pass preliminary control are evaluated by Editor-in-Chief for their originality, methodology, the importance of the subject covered and compliance with the journal scope. For the manuscripts that are considered for publication an unbiased, double-blind peer-review process is initiated by associate editors. The manuscripts are sent to two or more independent peer reviewers who are expert in the field. The reviewers are asked to provide detailed,

constructive comments that will help the editor to make a decision, and the authors to improve their manuscript. The requested changes are made by the authors using the online manuscript submission system. Associate editor sends the original manuscript, the revised manuscript and all the reviewers' comments together with his recommendation to the Editor-in-Chief.

Editor-in-Chief reviews the manuscript and makes the final decision. He might accept the manuscript, reject the manuscript or request further revisions.

In case of manuscripts submitted by editors or by the editorial board members of the journal, the editorial board will invite an external, independent editor to manage the evaluation processes of these manuscripts. The reviewers



INTERNATIONAL JOURNAL OF MEDICAL BIOCHEMISTRY

are required to notify the editor of any kinds of conflicts of interest with respect to the research, the authors and/or the research funders. They must ensure that all the information related to submitted manuscripts is kept as confidential and must report to the Editor-in-Chief if they are aware of copyright infringement and plagiarism on the author's side.

International Journal of Medical Biochemistry takes into consideration the Committee on Publication Ethics - COPE guidelines and flow charts for peer review process.

PUBLISHING FEE

Submission Fee

As of January 1, 2026, authors are required to pay a submission fee of USD 120 (including VAT) at the time of manuscript submission to the International Journal of Medical Biochemistry (IJMB). The submission process will not be initiated for manuscripts for which the submission fee has not been paid. This fee is charged to contribute to the costs of editorial handling. The submission fee is applied during the manuscript submission process and is non-refundable, regardless of the final editorial decision (acceptance or rejection). Kindly note that any bank transaction fees must be covered by the author.

Article Processing Charge (APC)

As of January 1, 2026, manuscripts submitted to the International Journal of Medical Biochemistry (IJMB) that are accepted for publication will be subject to an Article Processing Charge (APC) of USD 120 (including VAT). The publication process will not proceed unless the APC has been fully paid. This fee is collected to support the production, and publication of the accepted manuscripts. Kindly note that any bank transaction fees must be covered by the author.

Additional Information

Authors based in Türkiye may pay the above fees in Turkish Lira (TRY), calculated based on the USD-TRY exchange rate applicable at the time of payment. For any questions regarding payments, please contact us at kare@karepb.com.

Bank Information

KARE MEDYA LİMİTED ŞİRKETİ (Türk Yazarlar İçin TL Hesabı)

IBAN : TR92 0006 7010 0000 0026 2955 30

Account Holder: KARE MEDYA LIMITED SİRKETİ

Account Number: 26297654

IBAN: TR71 0006 7010 0000 0026 2976 54

SWIFT CODE: YAPITRISXXX

Address: Sahrayıcedit, Atatürk Cad. No: 43A, Kadıköy, İstanbul, Türkiye

PREPARATION AND SUBMISSION OF MANUSCRIPTS

Manuscript files should be prepared with Microsoft Office Word. The online submission system will direct authors during all stages of submission and provide necessary support for accelerating the submission process. A list of the files that should be supplied through the online submission system is provided below.

- Title Page
- Main Text
- Tables, Graphs and Figures
- Copyright Transfer Form
- Author Contribution Form

ICMJE Uniform Disclosure Form for Potential Conflicts of Interest.

MANUSCRIPT TYPES

Contributions may be in the form of clinical and basic Original Research articles, Reviews, **Short communications**, Case reports, Letters to the Editor, **Opinion papers**, Technical notes and **Editorial Comment**. The Journal also publishes brief reports on original studies or evaluations, book reviews, and proceedings of scientific meetings.

Original Research: Consists of Title, Authors, their addresses, Abstract, Key Words, Introduction, Materials and Methods, Results, Discussion, Ethical Considerations (if necessary), Acknowledgements, Conflict of Interest, References, Figure Legends, Figures, and Tables. It should not exceed 5000 words excluding the references. The manuscript should have no more than 50 references and a total of 6 tables and/or figures.

Review: Reviews prepared by authors with extensive knowledge on a particular field, which has been reflected in international literature by a high number of publications and citations, are evaluated. The authors may be invited by the Editor-in-Chief. A review should be prepared in the format describing, discussing and evaluating the current level of knowledge or topic that is to be used in the clinical practice and it should guide further studies. A review article consists of Title, Authors, their addresses, Abstract, Key Words, Introduction, Main Sections under headings written in bold and sentence case, Subsections (if any) under headings written in italic and numbered consecutively with Arabic numerals, Conclusion, Acknowledgements, Conflict of Interest, References, Figure Legends, Figures, and Tables. The manuscript should have no more than 75 references and a total of 6 tables and/or figures. Supplemental data are permitted for Review articles. Abstract should not include subheadings and should be limited to 300 words. Keywords section should contain a minimum of three and a maximum of six items in accordance with Medical Subject Headings (MeSH) terms prepared by the National Library of Medicine (NLM) and should be provided just below the abstract. Main Text should include Introduction, other subheadings and Conclusion sections and should be limited to 5000 words excluding the references. Tables, Figures and Images should be provided after the reference list according to their order of appearance in the text.

Note: The originality of the visuals included in the reviews should be assured by submission of an accompanying letter by the authors. Appropriate citation should be done for the visuals adapted from previously published sources, in accordance with the original versions of the printed or electronic copies. The written permission obtained from the copyright holder (publisher, journal or authors) should be sent to the Editor-in-Chief's Office.

Short communications: These manuscripts are intended to concise and quick publishing of a new finding. Publishing of research articles under this concept is decided by the Editorial Board, providing that the authors adhere to the publishing format. The general format of this type of manuscript is similar to that of research articles except the word and reference limitations. For the short communications, main text should not exceed 1,500 words and number of references should not exceed 15.

Case Report: Since a limited number of case reports is published, only reports which are related to rare cases and conditions that constitute challenges in diagnosis and treatment, offer new methods or suggest knowledge not included in books, and are interesting and educational are accepted for publication. A case presentation consists of Title, Authors, their addresses, Abstract, Key Words, Introduction, Patients and Methods, Results, Discussion, Conclusion, Ethical Considerations, Acknowledgements, Conflict of Interest, References, Figure Legends, Figures, and Tables. Main Text should not exceed 1500 words excluding the references. The reference list should follow the main text and the number of references should be limited to 15. Tables, Figures and Images should be provided after the reference list according to their order of appearance in the text and should be limited to two.



Letter to the Editor: Letters to the Editor aim to discuss the importance of a manuscript previously published in the journal. This type of manuscripts should also include a comment on the published manuscript. Moreover, articles on topics of interest to readers within the scope of the journal, especially on educational issues, can be published in the format of a Letter to the Editor. It consists of title, main text and references sections. **Abstract, Keywords, Tables, Figures and Images, and other visuals are not included.** Main Text should not include subheadings and it should be limited to 500 words. The reference list should follow the main text and the number of references should be limited to five. The volume, year, issue, page numbers, authors' names and title of the manuscript should be clearly stated, included in the list of references and cited within the text.

Opinion papers: Opinions on the topics within the scope of the journal that are prepared by the experts are published in this section. An opinion consists of Title, Authors, their addresses, Abstract, Key Words, Introduction, Discussion, Conclusion, Ethical Considerations (if necessary), Acknowledgements, Conflict Of Interest, References, Figure Legends, Figures, and Tables. Editorial Board decides the eligibility of an opinion with respect to its concept and language. For the opinions, main text should not exceed 1,500 words and number of references should not exceed 15.

Technical reports: Manuscripts on development and application of new methodologies are published in this category. A technical report consists of Title, Authors, their addresses, Abstract, Key Words, Introduction including the main aspects of the method involved, Materials and Methods, Results, Discussion, Ethical Considerations (if necessary), Acknowledgements, Conflict of Interest, References, Figure Legends, Figures, and Tables. Data analysis should be presented in Materials and Methods. Clinical technical reports should include and discuss the clinical significance of values and their deviations. For the technical reports, main text should not exceed 1,500 words and number of references should not exceed 10.

Editorial Comment: Authors are selected and invited by the Editor-in-Chief. This type of manuscript aims at providing a brief commentary on an article published in the journal by a researcher who is an authority in the relevant field or by the reviewer of the article. It should consist of title, main text and references sections. Main Text should not include subheadings and should be limited to 500 words. The reference list should follow the main text and the number of references should be limited by 15. Abstract, Keywords, Tables, Figures, Images and other visuals are not included in editorial comments.

PREPARATION OF MANUSCRIPTS

Manuscripts should be prepared according to the above mentioned word and reference limitations and other related information. Language: Manuscripts should be written in clear and concise English. Please have your text proofread by a native English speaker before you submit for consideration.

The manuscripts submitted to our journal are classified and evaluated according to the manuscript types stated below. During preliminary evaluation, the editors assess whether a manuscript's format and subheadings are prepared in accordance with the journal's guidelines. Therefore, it is important that authors check the accuracy of the main text in terms of the following.

Title Page: Title should be concise and informative and reflect the content of the manuscript and should not exceed 15 words. Avoid abbreviations and formulae where possible. It should be written in sentence case; that is, first letter of the initial word should be written in capital letter and rest of the title should be typed with lower case letters except proper nouns and abbreviations.

Information about the authors and their institutions should not be included in the main text, tables and figures. Since submitted manuscripts are evaluated by the reviewers through the online system, personal

identification is excluded in the interests of unbiased interpretation. Thus, only information about the manuscript as specified below should be included on the title page. For each type of manuscript, it is mandatory to upload a title page as a separate Microsoft Word document through the online submission system. The title page should include the names of the authors with their latest academic degrees, and the name of the department and institution, city and country where the study was conducted. If the study was conducted in several institutions, the affiliation of each author must be specified with symbols. The first letters of authors' names and surnames should begin with capital letter and the rest should be written in lower case letters. If there is more than one author, corresponding author should be indicated.

The correspondence address should contain the full name of the corresponding author, postal and e-mail addresses, phone and fax numbers. If the content of the manuscript has been presented before, the name, date and place of the meeting must be noted. **Disclosure of conflict of interest,** institutional and financial support, author contributions and acknowledgments should be included on the title page.

Structured Abstract: It should be structured with Objective, Methods, Results and Conclusion subheadings and should be limited to 300 words. The abstract should state briefly the purpose of the research, the principal results and major conclusions. An abstract is often presented separately from the article, so it must be able to stand alone. For this reason, References should be avoided. Also, non-standard or uncommon abbreviations should be avoided, but if essential they must be defined at their first mention in the abstract itself. Level of evidence has to be stated at the end of the abstract as a separate paragraph.

Keywords: This section should contain a minimum of three and a maximum of six items in accordance with Medical Subject Headings (MeSH) terms prepared by the National Library of Medicine (NLM) and should be placed just below the abstract.

Main Text: Text should be prepared with MS Word document. All text should be written with Arial font type at 12 font size and double spaced on A4 sized paper with 1 cm right margin and 3 cm other margins. Header and/or footer should not be inserted in the text except that the pages should be numbered with Arabic numerals. Main text should consist of title, introduction, materials and methods, results and discussion.

Introduction: State the objectives of the work and provide an adequate background, avoiding a detailed literature survey or a summary of the results. The unnecessary details and excessive references should be avoided. Hypothesis and the aim of the study should be stated in the introduction.

Material and methods: Provide sufficient detail to allow the work to be reproduced. Methods already published should be indicated by a reference; only relevant modifications should be described. Design and place of the study, subjects/patients, if applicable, information on the treatment procedures, statistical methods and information on the adherence to the ethical rules should be indicated.

Statistical analysis should be conducted in accordance with the guidelines on reporting statistical data in medical journals. The software used for statistical analysis must be described. The outcomes of statistical analyses and interpretation of the results must be in evidence-based scientific language. Statistics- statistical methods should be explained in detail in the Materials and Methods so that if original data is given, the results can be verified. If possible, quantitative results should be given and appropriate indicators should be used to indicate measurement error or uncertainty.

Results: Results should be clear and concise. Results must be concise and include figures and tables. Descriptive statistics should be compatible with the nature of data and statistical analyses used. The graphs should be prepared to reflect the important features of data. Please avoid excessive figures and tables.



Tables, Graphs and Figures

- Tables, Graphs, Figures and other visuals should be numbered in the order of their citation within the text and names of patients, doctors and institutions should not be disclosed.
- Tables should be prepared in a Microsoft Office Word document using the command 'Insert Table' and inserted at the end of the references in the main text.
- Tables should not be submitted in JPEG, TIFF or other visual formats. For microscopic images, the magnification ratio and staining technique used should be specified in addition to figure legends.
- Roman numbers should be avoided in tables and figures within the text and their titles. Each table should be clearly labeled and numbered consecutively with Arabic numerals (Table 1, Table 2, etc.). The abbreviations used must be defined. As with the text, reporting of concentration units consists of analyte concentrations expressed in the text in the traditional mass unit (mg/dL, ng/mL, and so forth) followed by the SI unit in parentheses
- The thickness of the lines in graphs should be sufficient to minimize loss of quality if size reduction is needed during the printing process. The width of the graphs should be 9 cm or 18 cm. Drawings should be performed by professionals. No grey colours should be used.
- Abbreviations should be explained in alphabetical order at the bottom of the tables, graphs and figures.
- Explanations should be at the bottom of the table as footnotes.
- Tables should not be divided into parts. Separate parts (Table 1A, Table 1B, etc.) will be considered separate tables and will count toward your overall table/figure count.
- Table captions should be limited to 60 words.
- Tables should be easily understandable and should not repeat the data in the main text.
- In addition to the pictures included in case reports and original images, video and movie images are published on the journal's website. These images should be prepared in MPEG format with a maximum size of 2 MB. They should be submitted to the journal with the manuscript documents. The names of patients, doctors, institutions and places should be omitted from all documents.
- All figures must be uploaded separately as image files in Tagged Image File Format (.TIFF) or Encapsulated Postscript (.EPS). Microsoft Power Point (.PPT) is also acceptable; however, the graphics must contain embedded fonts with one image per slide, one slide per file. Each image should have a resolution of 600 dots-per-inch (dpi).
- Indicate each footnote in a table with a superscript lowercase letter.

Discussion: This should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

Conclusions: The main conclusions of the study may be presented in a short Conclusions section, which may stand alone or form a subsection of a Discussion or Results and Discussion section.

Appendices: If there is more than one appendix, they should be identified as A, B, etc. Formulae and equations in appendices should be given separate numbering: Eq. (A.1), Eq. (A.2), etc.; in a subsequent appendix, Eq. (B.1) and so on. Similarly for tables and figures: Table A.1; Fig. A.1, etc.

Abbreviations: Define abbreviations that are not standard in this field in a footnote to be placed on the first page of the article. Such abbreviations that are unavoidable in the abstract must be defined at their first mention there, as well as in the footnote. Ensure consistency of abbreviations throughout the article.

Acknowledgements: Collate acknowledgements in a separate section at the end of the article before the references and do not, therefore, include

them on the title page, as a footnote to the title or otherwise. Financial and/or technical (running of tests, evaluation of results, providing language help, writing assistance or proof reading the article, etc.) supports of the study should be mentioned following Ethical Considerations including the project number. If the study originates from a thesis or has been previously presented in a meeting, this should also be indicated in this section. Contribution of each author should be declared in this section.

Nomenclature and units: Follow internationally accepted rules and conventions: use the international system of units (SI). If other quantities are mentioned, give their equivalent in SI. Authors wishing to present a table of nomenclature should do so on the second page of their manuscript. You are urged to consult IUB: Biochemical Nomenclature and Related Documents for further information.

Footnotes: It should be used sparingly. Number them consecutively throughout the article, using superscript Arabic numbers. Many word processors build footnotes into the text, and this feature may be used. Should this not be the case, indicate the position of footnotes in the text and present the footnotes themselves separately at the end of the article. Do not include footnotes in the Reference list.

Math formulae: Please submit math equations as editable text and not as images. Present simple formulae in line with normal text where possible and use the solidus (/) instead of a horizontal line for small fractional terms, e.g., X/Y. In principle, variables are to be presented in italics. Powers of e are often more conveniently denoted by exp. Number consecutively any equations that have to be displayed separately from the text (if referred to explicitly in the text).

Special Terms and Conditions: For double-blind peer-review process, the names of the corresponding author and other authors, their affiliations and any information on the study centres should not be included in any part of the submitted manuscripts and images, except the Title Page. This information should be added to the relevant section of the online submission system and included in the Title Page.

Pharmaceutical products should be written with their generic names and brand and company names, city and country should be specified for medical equipment and devices.

For centrifugation, angular velocity (xg) should be given. Period should be used to show decimals.

Enzymes: Full name of an enzyme and its International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme Commission number (EC number) should be given when that enzyme was first mentioned in the text. Its abbreviation can be used after that. For detailed information, authors can refer to the Enzyme Nomenclature of IUBMB (www.chem.qmul.ac.uk/iubmb/enzyme). Enzyme activity should be given in terms of U/L in serum or plasma and U/mg protein in tissues. Unit (U) should be given as reaction rate triggered under defined conditions. SI unit for enzyme activity is defined as katal which is the moles of substrate converted to product per second. However, it can also be defined as $\mu\text{mol}/\text{sec}$ or $\mu\text{mol}/\text{min}$ as long as clear and precise definitions are given.

References

- References should start on a separate page. Published manuscripts and manuscripts that have been accepted and are pending publication should be cited in the reference list. Acceptance letter of the manuscripts that are pending publication should be submitted with the manuscript. Unpublished results should be written in text in parenthesis as "(unpublished result)". If such a work is cited, permission from the responsible researcher should be obtained and submitted with the manuscript.
- The reference list containing all authors mentioned in the text, should be arranged in the order of appearance and numbered accordingly by Arabic numerals in square brackets in line with the text. When more than one reference is cited consecutively, not all of the reference



numbers should be written separated by commas; instead the numbers of first and the last reference cited for that section should be written with a dash in the middle (i.e.: writing as "10, 11, 12, 13" is wrong, but "10- 13" is correct). Citation of a reference as 'in press' implies that the item has been accepted for publication.

- References that are inaccessible and not indexed in any database should not be cited.
- The actual authors can be referred to, but the reference number(s) must always be given. Example: '..... as demonstrated [7,8]. Dildar and Murat [9] obtained a different result'
- Authors' names are inverted (last name, first/second initial). Do not add periods or commas within an individual author name; however, separate author names with a comma and end the author list with a period (Konukoglu D, Ilhan N, Orem A)
- The Journal names should be abbreviated as indicated at PubMed.
- For references with six and fewer authors, all authors should be listed. For references with more than six authors, the first six authors should be listed, followed by 'et al'.
- Do not add a period after the journal abbreviation, but continue with a space followed by the year.
- The year should be followed by a semicolon and then the volume number, which is followed by a colon and then the page numbers. Delete redundant numbers, for example 2017;12:101-15.
- Do not include the months in parentheses; this information is not needed.
- Use inclusive page numbers for articles and book chapters.

Data References: This journal encourages you to cite underlying or relevant datasets in your manuscript by citing them in your text and including a data reference in your Reference List. Data references should include the following elements: author name(s), dataset title, data repository, version (where available), year, and global persistent identifier. Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Web References: As a minimum, the full URL should be given and the date when the reference was last accessed. Any further information, if known (DOI, author names, dates, reference to a source publication, etc.), should also be given. Web references can be listed separately (e.g., after the reference list) under a different heading if desired, or can be included in the reference list.

References in a Special Issue: Please ensure that the words 'this issue' are added to any references in the list (and any citations in the text) to other articles in the same Special Issue. The style and punctuation of the references should be formatted as in the following examples.

Technical Reports: Surname(s) and initial(s) of the author(s), title of the technical report, name and place of the publisher, report number, if applicable, publication year. The style and punctuation of the references should be formatted as in the following examples. Journal Article

- Journal article with six or fewer authors: Bozluolcay M, Andican G, Firtina S, Erkol G, Konukoglu D. Inflammatory hypothesis as a link between Alzheimer's disease and diabetes mellitus. *Geriatr Gerontol Int* 2016 16(10):1161-66
- Journal article with more than six authors: Koca SS, Kara M, Özgen M, Dayanan R, Demir CF, Aksoy K, et al. Low prevalence of obesity in Behçet's disease is associated with high obestatin level. *Eur J Rheumatol* 2017; 4 (2):113-17.
- Abstract: Hortin GL, King C, Kopp J. Quantification of rhesus monkey albumin with assays for human microalbumin [Abstract]. *Clin Chem* 2000;46:A140-
- Editorial: Demers LM. New biochemical marker for bone disease: is it a breakthrough? [Editorial]. *Clin Chem* 1992;38:2169-70.

- Letter to the Editor: Davey L, Naidoo L. Urinary screen for acetaminophen (paracetamol) in the presence of N-acetylcysteine [Letter]. *Clin Chem* 1993;39:2348-9.
- Book Chapter: Rifai N, Warnick GR. Lipids, lipoproteins, apolipoproteins, and other cardiovascular risk factors. In: Burtis CA, Ashwood ER, Bruns DE, editors. *Tietz textbook of clinical chemistry and molecular diagnostics*. 4th Ed. St. Louis (MO): Elsevier Saunders; 2006. p. 903-81.
- Book: Personal Author Harrell FE Jr. Regression modeling strategies. New York (NY): Springer; 2001. Bailer JC III, Mosteller F, editors. *Medical uses of statistics*. 2nd Ed. Boston (MA): NEJM Books; 1992:449 p.
- Book with Single Author Cohn PF. Silent myocardial ischemia and infarction. 3rd ed. New York: Marcel Dekker; 1993.
- Editor(s) as author: Norman LJ, Redfern SJ, editors. *Mental health care for elderly people*. New York: Churchill Livingstone; 1996
- Technical Reports Tschantz BA, Moran B. Modeling of the hydrologic transport of mercury in the Upper East Fork Poplar Creek (UEFPC) watershed. Technical Report for Lockheed Martin Energy Systems: Bethesda, MD, September 2004.
- Supplement Castelli WP. Lipids, risk factors and ischaemic heart disease. *Atherosclerosis* 1996;124 Suppl:S1-9.
- Epub ahead of print Milbury CA, Li J, Makrigiorgos GM. PCR-based methods for the enrichment of minority alleles and mutations. [Epub ahead of print] *Clin Chem* February 6, 2009 as doi:10.1373/clinchem.2008.113035.
- Internet Source: American Association for Clinical Chemistry. AACC continuing education. <http://www.aacc.org/education-and-career/continuing-education>(Accessed June 2016).

Reviewer Suggestion: Authors may provide the names of two proposed reviewers in the "Comments to Editorial Office" screen during the submission process.

EVALUATION AND PUBLICATION

The main text submitted to the journal must be in English. Manuscripts written in other languages are not accepted. Citation potential is higher for manuscripts in English.

The Editor-in-Chief's Office checks the conformity of the manuscript with the journal's general guidelines before sending it to associate editors and reviewers. Any manuscripts not prepared in accordance with the journal's guidelines will be returned for revision. The evaluation period is limited to 30 days. If revision is requested, authors should submit their revised manuscripts within 45 days. Manuscripts and revisions should be submitted through the online manuscript submission system at the website <http://www.internationalbiochemistry.com>. Manuscripts sent by e-mail will not be accepted.

Manuscripts are evaluated in accordance with the principles of double-blind peer review. Of the submitted manuscripts, those considered to be suitable are subjected to evaluation in terms of style, format and in terms of their content by the Editor-in-Chief. Manuscripts considered to be scientifically adequate are assigned to three reviewers. These reviewers are independent experts and members of the editorial board who have published internationally on the topic of the manuscript. Research articles, systematic reviews and meta-analyses are also evaluated by the Editor-in-Chief's statistician counsellors in addition to peer review. When needed, the manuscripts are scanned by the Editor-in-Chief's Office using the iThenticate program for determination of plagiarism and non-ethical situations. Also, all manuscripts are reviewed and edited for compliance with the rules of English grammar. All contributing authors of a manuscript accepted for publication are deemed to have accepted the right of editors to make corrections provided that no fundamental change is made to the basic meaning of the original text. Manuscripts not accorded priority for publication by the Editor-in-Chief, Editor, Associate Editors and Referees are rejected.



INTERNATIONAL JOURNAL OF MEDICAL BIOCHEMISTRY

The retraction requests of the manuscripts close to the final decision in evaluation process, without any reasonable explanation (not contrary to the Publication Ethics) is considered in the scope of rejection decision. Manuscripts that are not revised and resubmitted within the specified time and fail to notify the journal accordingly will be rejected.

Once a manuscript is accepted for publication, addition to the author list, removal from the author list and order change cannot be effected.

Manuscripts accepted for publication are prepared as 'Epub ahead of print articles' and published with an assigned DOI on the journal website <http://www.internationalbiochemistry.com>.

Information on the latest status of manuscripts submitted for evaluation and other information about the journal are available on the website <http://www.internationalbiochemistry.com>. Contact details for the Editor-in-Chief's Office and the publisher are given below.

Online proof correction: Corresponding authors will receive an e-mail with a link to our online proofing system, allowing annotation and correction of proofs online. The environment is similar to MS Word: in addition to editing text, you can also comment on figures/tables and answer questions from the Copy Editor. Web-based proofing provides a faster and less error-prone process by allowing you to directly type your corrections, eliminating the potential introduction of errors.

If preferred, you can still choose to annotate and upload your edits on the PDF version. All instructions for proofing will be given in the e-mail we send to authors, including alternative methods to the online version and PDF.

We will do everything possible to get your article published quickly and accurately. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. It is important to ensure that

all corrections are sent back to us in one communication. Please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility.

Submission checklist: You can use this list to carry out a final check of your submission before you send it to the journal for review. Please check the relevant section in this Guide for Authors for more details.

Ensure that the following items are present: One author has been designated as the corresponding author with contact details:

- ORCID id for all authors
- E-mail address
- Full postal address
- All necessary files have been uploaded: Manuscript: Include keywords
- All figures (include relevant captions)
- All tables (including titles, description, footnotes)
- Ensure all figure and table citations in the text match the files provided • Indicate clearly if color should be used for any figures in print Graphical Abstracts / Highlights files (where applicable) Supplemental files (where applicable) Further considerations
- Manuscript has been 'spell checked' and 'grammar checked'
- All references mentioned in the Reference List are cited in the text, and vice versa
- Permission has been obtained for use of copyrighted material from other sources (including the Internet)
- A competing interests statement is provided, even if the authors have no competing interests to declare
- Journal policies detailed in this guide have been reviewed
- Referee suggestions and contact details provided, based on journal requirements.

ETHICS AND POLICIES

ADVERTISEMENT POLICY

The International Journal of Medical Biochemistry does not accept advertisements; the journal is sponsored by the Association of Clinical Biochemistry Specialists Turkey. The Journal has no direct marketing policy.

ARCHIVING POLICY

The content published by the International Journal of Medical Biochemistry is electronically preserved by using internet archive. <https://internationalbiochemistry.com/jvi.aspx?pdid=ijmb&plng=eng&list=pub>

AUTHORSHIP POLICY

All statements and opinions expressed in the manuscripts published in the International Journal of Medical Biochemistry reflect the views of the author(s).

Each individual listed as an author should fulfill the authorship criteria recommended by the International Committee of Medical Journal Editors (ICMJE). The ICMJE recommends that authorship should be based on the following 4 criteria:

Substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of data for the work; AND

Drafting the work or revising it critically for important intellectual content; AND

Final approval of the version to be published; AND

Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

In addition to being accountable for their own work, authors should have confidence in the integrity of the contributions of their co-authors and each author should be able to identify which co-authors are responsible for other parts of the work.

All of those designated as authors should meet all four criteria for authorship, and all who meet the four criteria should be identified as authors. Those who provided a contribution but do not meet all four criteria should be recognized separately on the title page and in the Acknowledgements section at the conclusion of the manuscript.

The International Journal of Medical Biochemistry requires that corresponding authors submit a signed and scanned version of the authorship contribution form (available for download through) during the initial submission process in order to appropriately indicate and observe authorship rights and to prevent ghost or honorary authorship. Please note that the list of authors on the final manuscript will be presented in the order provided on this form. If the editorial board suspects a case of "gift authorship," the submission will be rejected without further review. As part of the submission of the manuscript, the corresponding author should also send a short statement declaring that they accept all responsibility for authorship during the submission and review stages of the manuscript. Requests for authorship changes (additions, deletions, or order changes) after submission will be handled in accordance with COPE recommendations and require the agreement of all listed authors.

COMPLAINT AND APPEAL POLICY

Appeal and complaint cases are handled within the scope of COPE guidelines by the Editorial Board of the journal. Appeals should be based on the scientific content of the manuscript. The final decision on the appeal and complaint is made by Editor in Chief. An Ombudsperson or the Ethical



Editor is assigned to resolve cases that cannot be resolved internally. Authors should get in contact with the Editor in Chief regarding their appeals and complaints via e-mail at kare@karepb.com

CONFIDENTIALITY

All submitted manuscripts are considered confidential material. International Journal of Medical Biochemistry will not disclose the content of any submission to individuals other than those directly involved in the editorial and publication process, including editorial staff, corresponding authors, potential and invited reviewers, and editors.

In cases where misconduct is suspected, the journal reserves the right to share the manuscript with members of relevant ethics committees or with institutions and organizations as necessary to facilitate the investigation and resolution of the alleged misconduct.

CORRECTIONS POLICY

If the editors or publisher learn from a third party that a published work contains a material error or inaccuracy, the authors must promptly correct

or retract the article or provide the journal editors with evidence of the accuracy of the article.

ETHICS POLICY

The Editorial Board of the International Journal of Medical Biochemistry and the Publisher adheres to the principles of the International Council of Medical Journal Editors (ICMJE), the World Association of Medical Editors (WAME), the Council of Science Editors (CSE), the Committee on Publication Ethics (COPE), the US National Library of Medicine (NLM), the World Medical Association (WMA) and the European Association of Science Editors (EASE).

The manuscripts should be prepared in accordance with ICMJE-Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals (updated in May 2022 - <https://www.icmje.org/recommendations/>).

Authors are required to prepare manuscripts in accordance with the international guidelines* below.

Randomized Controlled Trial **	CONSORT (Consolidated Standards of Reporting Trials)
Non-Randomized Trial **	TREND (Transparent Reporting of Evaluations with Non-randomised Designs)
Trial Protocol	SPIRIT (Standard Protocol Items Recommendations for Interventional Trials)
Observational Epidemiologic Studies (cohort, case-control, cross-sectional)	STROBE (the STrengthening the Reporting of OBServational studies in Epidemiology)
Systematic Reviews and Meta-Analyses	PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses)
Systematic Reviews and Meta-Analyses Protocol	PRISMA-P (the Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Protocol)
Qualitative Research	-SRQR (the Standards for Reporting Qualitative Research) -COREQ (Consolidated criteria for REporting Qualitative research: interviews and focus groups)
Methodological Studies (Translating and Adapting Tests)	ITC (International Test Commission) Guidelines for Translating and Adapting Tests
Methodological Studies (Developing Tests)	COSMIN (Consensus based Standarts fort the selection of Health Measurement Instruments- Study Design for Patient-reported outcome measurement instruments)
Case Report	CARE (CAse REport)
* Enhancing the QUALity and Transparency of Health Research (equator network)	
** The International Journal of Medical Biochemistry encourages the registration of all clinical trials (randomized and non-randomized) via ClinicalTrials.gov or one of the registries of the WHO's International Clinical Trials Registry Platform (ICTRP). The name of the trial registry and the registration number together should be provided at the end of the abstract.	

In accordance with the journal's policy, an approval of research protocols by an ethics committee in accordance with international agreements "WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects (last updated: October 2013, Fortaleza, Brazil)", "Guide for the care and use of laboratory animals (8th edition, 2011)" and/or "International Guiding Principles for Biomedical Research Involving Animals (2012)" is required for all research studies. If the submitted manuscript does not include ethics committee approval, it will be reviewed according to COPE's guideline (Guidance for Editors: Research, Audit and Service Evaluations). If the study should have ethical approval, authors will be asked to provide ethical approval in order to proceed the review process. If they cannot provide ethical approval, their manuscript will be rejected and also their institutions and when needed, the related bodies in their country will be informed that such studies must have ethics committee approval. If they provide approval, review of the manuscript will continue.

If the study does not need ethics committee approval after the editorial board's review, the authors will be asked to provide an ethics committee approval or a document given by a related independent committee that indicates the study does not need ethics committee approval according to the research integrity rules in their country. If the authors provide either an approval or a document showing that ethics approval is not needed, the review process can be continued. If the authors cannot provide either documents, the manuscript may be rejected.

For articles concerning experimental research on humans, a statement should be included that shows informed consent of patients and volunteers was obtained following a detailed explanation of the procedures that they may undergo. The journal may request a copy of the Ethics Committee Approval received from the relevant authority. Informed consent must also be obtained for case reports and clinical images.

Studies using human or animal subjects should be approved by the appropriate institutional and local Ministry of Health ethics committees. Ethics approval of research protocols in accordance with international agreements is required for experimental, clinical, and drug studies, as well as for some case reports. Ethics committee reports or an equivalent official document may be requested from the authors. For manuscripts involving experimental research on humans, a statement should be included that shows that written, informed consent of patients and volunteers was obtained. For studies carried out on animals, the measures taken to prevent pain and suffering of the animals should be stated clearly. A statement regarding patient consent, and the name of the ethics committee, the ethics committee approval date, and number should be stated in the Materials and Methods section of the manuscript. It is the authors' responsibility to carefully protect patients' anonymity.

RESEARCH ETHICS FOR VULNERABLE POPULATIONS

At the International Journal of Medical Biochemistry, we are committed to upholding the highest ethical standards in all research involving human



INTERNATIONAL JOURNAL OF MEDICAL BIOCHEMISTRY

participants, especially vulnerable populations such as children. In line with our dedication to responsible and respectful research practices, we have established the following guidelines to ensure the protection and ethical treatment of these groups:

CONSENT REQUIREMENTS FOR CHILDREN

Parental/Guardian Consent: For all research involving children under the age of 18, written informed consent must be obtained from a parent or legal guardian. This consent must be informed, voluntary, and documented.

Assent from Children: In addition to parental consent, researchers are required to obtain assent from children who are capable of forming an opinion and making a decision regarding their participation in the study. This process must be age-appropriate and must respect the child's level of understanding and autonomy.

Privacy and Confidentiality: Extra precautions will be taken to protect the privacy and confidentiality of child participants. This includes using pseudonyms, removing identifiable details from published data, and securely storing data.

Ethical Review: All studies involving children must undergo a rigorous ethical review process to ensure that the research is justified, and the potential benefits outweigh any risks. The ethical review will also ensure that the study adheres to the principles of beneficence, non-maleficence, and justice.

OVERSIGHT AND MONITORING

To ensure adherence to these ethical guidelines, the International Journal of Medical Biochemistry requires that all studies involving vulnerable populations be reviewed and monitored by an Institutional Review Board (IRB) or an equivalent ethical oversight committee. This committee will oversee the study from its inception to its completion, ensuring continuous protection of the participants' rights and well-being.

For more details on our research ethics policies and procedures, or to report any concerns regarding the ethical conduct of a study published in our journal, please contact our ethics committee at kare@karepb.com.

FEE WAIVER POLICY

There is no fee waiver.

FUNDING SOURCES POLICY

All authors are required to declare what support they received to carry out their research. Declaring funding sources acknowledges funders' contributions, fulfills funding requirements, and promotes greater transparency in the research process.

Each author must individually declare all sources of funding received for the research submitted to the journal. This information includes the name of granting agencies, grant numbers, and a description of each funder's role. If the funder has played no role in the research, this must be stated as well.

Authors are not required to provide the complete list of every single grant that supports them if the grant is not related to the research published.

LICENSES AND COPYRIGHT POLICY

Authors publishing with the journal retain the copyright to their work licensed under the Creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0) and grant the Publisher non-exclusive commercial right to publish the work. CC BY-NC 4.0 license permits unrestricted, non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

OPEN ACCESS POLICY

The International Journal of Medical Biochemistry supports the Budapest Open Access Initiative statement of principles that promotes free access to research literature. The declaration defines open access to academic literature as free availability on the internet, permitting users to read, record, copy, print, search, or link to the full text, examine them for indexing, use them as data for software or other lawful purposes without

financial, legal, or technical barriers. Information sharing represents a public good, and is essential to the advancement of science. Therefore, articles published in this journal are available for use by researchers and other readers without permission from the author or the publisher provided that the author and the original source are cited. The articles in the International Journal of Medical Biochemistry are accessible through search engines, websites, blogs, and other digital platforms. Additional details on the Budapest Open Access Initiative and their guidelines are available at <https://www.budapestopenaccessinitiative.org>.

Open Access Policy is based on rules of Budapest Open Access Initiative the International Journal of Medical Biochemistry applies the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license articles we publish. If you submit your paper for publication, you agree to have the CC BY-NC 4.0 license applied to your work. Under this Open Access license, you as the author agree that anyone can copy, distribute or reuse the content of your article for non-commercial purposes for free as long as the author and original source are properly cited. The corresponding author must sign the Creative Commons License Agreement after their articles are accepted.

OPEN ACCESS STATEMENT

The journal is an open access journal and all content is freely available without charge to the user or his/her institution. Except for commercial purposes, users are allowed to read, download, copy, print, search, or link to the full texts of the articles in this journal without asking prior permission from the publisher or the author. This is in accordance with the BOAI definition of open access. The open access articles in the journal are licensed under the terms of The open access articles in the journal are licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) license.



PEER REVIEW POLICY

Only those manuscripts approved by its every individual author and that were not published before in or sent to another journal, are accepted for evaluation.

Submitted manuscripts that pass preliminary control are scanned for plagiarism using iThenticate software. After plagiarism check, the eligible ones are evaluated by Editor-in-Chief for their originality, methodology, the importance of the subject covered and compliance with the journal scope. Editor-in-Chief evaluates manuscripts for their scientific content without regard to ethnic origin, gender, sexual orientation, citizenship, religious belief or political philosophy of the authors and ensures a fair double-blind peer review of the selected manuscripts.

The selected manuscripts are sent to at least two national/international referees for evaluation and publication decision is given by Editor-in-Chief upon modification by the authors in accordance with the referees' claims.

Editor-in-Chief does not allow any conflicts of interest between the authors, editors and reviewers and is responsible for final decision for publication of the manuscripts in the Journal.

Reviewers' judgments must be objective. Reviewers' comments on the following aspects are expected while conducting the review.

- Does the manuscript contain new and significant information?
- Does the abstract clearly and accurately describe the content of the manuscript?
- Is the problem significant and concisely stated?
- Are the methods described comprehensively?
- Are the interpretations and conclusions justified by the results?
- Are adequate references made to other Works in the field?
- Is the language acceptable?



INTERNATIONAL JOURNAL OF MEDICAL BIOCHEMISTRY

Reviewers must ensure that all the information related to submitted manuscripts is kept as confidential and must report to the editor if they are aware of copyright infringement and plagiarism on the author's side.

A reviewer who feels unqualified to review the topic of a manuscript or knows that its prompt review will be impossible should notify the editor and excuse himself from the review process.

The editor informs the reviewers that the manuscripts are confidential information and that this is a privileged interaction. The reviewers and editorial board cannot discuss the manuscripts with other persons. The anonymity of the referees is important.

PLAGIARISM POLICY

All submissions are screened using similarity detection software at least two times: on submission and after completing revisions. In the event of alleged or suspected research misconduct, e.g., plagiarism, citation manipulation, or data falsification/fabrication, the editorial board will follow and act in accordance with COPE guidelines. Plagiarism, including self-plagiarism, that is identified at any stage will result in rejection of the manuscript.

PUBLICATION CHARGES POLICY

Submission Fee

As of January 1, 2026, authors are required to pay a submission fee of USD 120 (including VAT) at the time of manuscript submission to the International Journal of Medical Biochemistry (IJMB). The submission process will not be initiated for manuscripts for which the submission fee has not been paid. This fee is charged to contribute to the costs of editorial handling. The submission fee is applied during the manuscript submission process and is non-refundable, regardless of the final editorial decision (acceptance or rejection). Kindly note that any bank transaction fees must be covered by the author.

Article Processing Charge (APC)

As of January 1, 2026, manuscripts submitted to the International Journal of Medical Biochemistry (IJMB) that are accepted for publication will be subject to an Article Processing Charge (APC) of USD 120 (including VAT). The publication process will not proceed unless the APC has been fully paid. This fee is collected to support the production, and publication of the accepted manuscripts. Kindly note that any bank transaction fees must be covered by the author.

Additional Information

Authors based in Türkiye may pay the above fees in Turkish Lira (TRY), calculated based on the USD-TRY exchange rate applicable at the time of payment. For any questions regarding payments, please contact us at kare@karepb.com.

Bank Information

KARE MEDYA LİMİTED ŞİRKETİ (Türk Yazarlar İçin TL Hesabı)

IBAN : TR92 0006 7010 0000 0026 2955 30

Account Holder: KARE MEDYA LIMITED SİRKETİ

Account Number: 26297654

IBAN: TR71 0006 7010 0000 0026 2976 54

SWIFT CODE: YAPITRISXXX

Address: Sahrayıcedit, Atatürk Cad. No: 43A, Kadıköy, İstanbul, Türkiye

RETRACTION POLICY

The publisher will take all appropriate measures to modify the article in question, in close cooperation with the editors, in cases of alleged or proven scientific misconduct, fraudulent publication, or plagiarism. This includes the prompt publication of an erratum, disclosure, or retraction of the affected work in the most severe case. Together with the editors, the publisher will take reasonable steps to detect and prevent the publication of articles in which research misconduct occurs and will under no circumstances promote or knowingly allow such abuse to occur.

WITHDRAWAL POLICY

The International Journal of Medical Biochemistry is committed to providing high quality articles and uphold the publication ethics to advance the intellectual agenda of science. We expect our authors to comply with, best practice in publication ethics as well as in quality of their articles.

Withdrawal of a manuscript will be permitted only for the most compelling and unavoidable reasons. For withdrawal of a manuscript authors need to submit an "Article withdrawal Form", signed by all authors mentioning the reason for withdrawal to the Editorial Office. The form is available from the web page of the journal. Authors must not assume that their manuscript has been withdrawn until they have received appropriate notification to this effect from the editorial office.

In a case where a manuscript has taken more than five months' time for review process, that allows the author to withdraw manuscript.

Manuscript withdrawal penalty: After receiving the Article withdrawal Form, the International Journal of Medical Biochemistry Editorial Board will investigate the reason of withdrawal.

If the reason finds to be acceptable, the author is allowed to withdraw the manuscript without paying any withdrawal penalty. If not the International Journal of Medical Biochemistry will not accept any manuscripts from the same author for one year.

Important notes: Manuscripts may be withdrawn at any stage of review and publication process by submitting a request to the editorial office. Manuscript withdrawal will be permitted after submission only for the most compelling and unavoidable reasons.



CONTENTS VOLUME IX ISSUE 4 YEAR 2026

EDITORIAL

From measurement to meaning

Konukoglu D.....	XIII
------------------	------

ORIGINAL ARTICLES

Systemic inflammatory and immunonutritional profile in women with low anti-Müllerian hormone levels

Dundar Akin O, Misirlioglu NF, Uzun H.....	201
--	-----

Clinical and laboratory correlates of low fecal elastase-I levels: A retrospective analysis

Hacihasanoglu Cakmak N, Ulfer G, Yigit P, Yigitbasi T, Canbaloglu G.....	210
--	-----

Comparability of adrenocorticotrophic hormone measured in spray-dried K2-EDTA and spray-dried K3-EDTA tubes

Ustundag Y, Karaca MS, Huysal K, Erdogan F.....	217
---	-----

End-user verification of the Maccura F810 hematology analyzer: A comparison with Mindray BC-6200

Cinpolat HY, Sehitoglu Bingol MH.....	223
---------------------------------------	-----

Interpretation of heterophile-blocking tube results in thyroid function testing using analytical change limits: An exploratory study of clinically discordant cases

Turkal R, Yaman A, Cevlik T, Gozu H, Sirikci O, Haklar G.....	232
---	-----

Internal validation and explainability of machine learning models for abnormal liver biochemistry

Demir O, Katar M, Nasip OF.....	242
---------------------------------	-----

Specimen type as the dominant predictor of pre-analytical rejection in the emergency department: An explainable machine-learning (SHAP) analysis

Goren Y, Ozerol H.....	251
------------------------	-----



EDITORIAL

From measurement to meaning

The studies in this issue reflect the breadth of contemporary medical biochemistry: from biomarker associations and analytical performance to assay interference, specimen quality, and machine learning. Together, they emphasize that laboratory results gain clinical value only when their limitations and context are clearly understood.

The studies of anti-Müllerian hormone and fecal elastase-I explore associations between laboratory findings and clinical or nutritional features. The ACTH tube comparison and hematology analyzer verification address a complementary question: whether methods and specimen conditions provide results fit for clinical use. Their findings also show why performance should be assessed at relevant concentration ranges and medical decision points.

The thyroid interference study highlights the need to interpret laboratory results alongside the clinical picture. The machine-learning studies extend this principle to larger datasets: strong model performance is informative, but outcome definitions, specimen mix, validation, and clinical relevance must also be considered.

Across these topics, a common responsibility emerges for laboratory professionals: to produce reliable results, recognize uncertainty, and communicate evidence with appropriate caution. We hope this issue encourages continued collaboration between laboratory specialists, clinicians, and researchers to strengthen the contribution of medical biochemistry to patient care.

Prof. Dildar Konukoglu, MD.

Editor-in-Chief



Research Article

Systemic inflammatory and immunonutritional profile in women with low anti-Müllerian hormone levels

Oznur Dunder Akin¹, Naile Fevziye Misirlioglu², Hafize Uzun³

¹Department of Obstetrics and Gynecology, Istanbul Atlas University Faculty of Medicine, Istanbul, Türkiye

²Department of Biochemistry, Istanbul Nisantasi University Faculty of Medicine, Istanbul, Türkiye

³Department of Biochemistry, Istanbul Atlas University Faculty of Medicine, Istanbul, Türkiye

Abstract

Objectives: Anti-Müllerian hormone (AMH) is widely used as a biomarker of ovarian reserve. Although systemic inflammation and nutritional status may be associated with ovarian function, the relationships of composite inflammatory and nutrition-related indices with low AMH remain incompletely characterized. This study aimed to evaluate systemic inflammatory and nutrition-related indices in women with low AMH levels.

Methods: A total of 150 women were included, comprising 75 women with low AMH and 75 with high AMH. The neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), systemic immune-inflammation index (SII), systemic inflammation response index (SIRI), aggregate index of systemic inflammation (AISI), prognostic nutritional index (PNI), and hemoglobin–albumin–lymphocyte–platelet (HALP) index were evaluated. Associations between continuous AMH concentrations and these indices were assessed using Spearman correlation analysis. Age- and body mass index (BMI)-adjusted logistic regression models were used to evaluate associations with low AMH.

Results: Women with low AMH were older than those with high AMH (33.0 [29.5–36.5] vs. 27.0 [24.0–31.0] years; $p < 0.001$), whereas BMI did not differ significantly between the groups. AISI, SIRI, PLR, NLR, and SII were significantly higher, whereas PNI and HALP were significantly lower, in the low AMH group, with all differences remaining significant after multiple-comparison correction. AMH was inversely correlated with AISI ($p = -0.282$), SIRI ($p = -0.266$), PLR ($p = -0.257$), NLR ($p = -0.314$), and SII ($p = -0.324$), and positively correlated with PNI ($p = 0.466$) and HALP ($p = 0.366$) (all $p \leq 0.001$). After adjustment for age and BMI, higher NLR was associated with greater odds of low AMH (adjusted OR=2.03, 95% CI: 1.34–3.06; $p < 0.001$), whereas higher PNI was associated with lower odds (adjusted OR=0.31, 95% CI: 0.19–0.51; $p < 0.001$).

Conclusion: Low AMH was associated with a profile of higher systemic inflammatory indices and lower nutrition-related indices. These associations remained significant after adjustment for age and BMI, supporting an association between low AMH and a distinct systemic inflammatory and immunonutritional profile. Further prospective, multicenter studies are needed to confirm these associations and clarify their biological and potential clinical significance.

Keywords: Anti-Müllerian hormone, HALP, neutrophil-to-lymphocyte ratio, ovarian reserve, prognostic nutritional index, systemic inflammation

How to cite this article: Dunder Akin O, Misirlioglu NF, Uzun H. Systemic inflammatory and immunonutritional profile in women with low anti-Müllerian hormone levels. Int J Med Biochem 2026;9(4):201–209.

Ovarian reserve represents the quantity of the remaining follicular pool and is a major determinant of female reproductive potential. Anti-Müllerian hormone (AMH), a glycoprotein produced predominantly by granulosa cells of preantral and small antral follicles, is widely used as a biomarker of

functional ovarian reserve because circulating AMH concentrations are closely related to the number of growing follicles [1, 2]. In contrast to several other reproductive hormones, AMH shows relatively limited intra-cycle variation and is clinically useful for assessing ovarian response, particularly in

Address for correspondence: Oznur Dunder Akin, MD. Department of Obstetrics and Gynecology, Istanbul Atlas University Faculty of Medicine, Istanbul, Türkiye

Phone: +90 532 740 58 72 **E-mail:** oznur.akin@atlas.edu.tr **ORCID:** 0000-0002-0205-9140

Submitted: September 03, 2026 **Revised:** September 16, 2026 **Accepted:** September 17, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



women undergoing assisted reproductive treatment [1]. Nevertheless, AMH primarily reflects follicle quantity rather than oocyte quality or the probability of spontaneous conception, and its concentration is strongly influenced by age and may also vary according to biological and clinical factors [1, 2].

Ovarian aging is a complex biological process characterized not only by progressive follicular depletion but also by alterations in the ovarian microenvironment. Accumulating evidence suggests that chronic low-grade inflammation, oxidative stress, mitochondrial dysfunction, cellular senescence, and immune dysregulation may contribute to age-related deterioration of ovarian function [3, 4]. Chronic inflammation may adversely affect folliculogenesis through changes in cytokine signaling, immune-cell activity, stromal fibrosis, oxidative stress, and follicular-cell survival, potentially linking systemic inflammatory status with ovarian reserve [3]. However, whether systemic inflammation is a cause, consequence, or merely a correlate of declining ovarian function remains unresolved, particularly in women without a defined inflammatory or reproductive disorder.

Several readily available hematological indices have therefore been investigated as surrogate markers of systemic inflammatory activity in reproductive medicine. The neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and systemic immune-inflammation index (SII) integrate different components of innate and adaptive immune responses and may provide broader information than individual blood-cell counts. Previous findings, however, have not been entirely consistent. Ilhan et al. [5] reported higher NLR values in women with premature ovarian insufficiency and demonstrated an inverse relationship between NLR and AMH. In contrast, a more recent study of women with diminished ovarian reserve found no significant association between NLR, PLR, or the red cell distribution width-to-platelet ratio and diminished ovarian reserve after considering other clinical factors [6]. More recently, a large retrospective cohort of 19,282 women undergoing assisted reproductive treatment showed that higher SII, PLR, and NLR were associated with lower AMH concentrations, lower antral follicle counts, and reduced ovarian response [7]. Collectively, these findings suggest a possible relationship between systemic inflammatory status and ovarian reserve while also highlighting the need for studies evaluating multiple inflammatory indices within the same population.

Beyond inflammation alone, nutritional and immunological status may also be relevant to reproductive physiology. The prognostic nutritional index (PNI), calculated from serum albumin concentration and lymphocyte count, is a simple composite measure integrating nutritional and immune-related information [8]. Similarly, the hemoglobin–albumin–lymphocyte–platelet (HALP) score incorporates hematological, nutritional, and inflammatory components and can be calculated from routinely available laboratory parameters [9]. Although PNI and HALP have been increasingly evaluated as immunonutritional markers in various clinical settings,

their relationship with ovarian reserve has received substantially less attention than conventional inflammatory indices. Moreover, the combined assessment of nutrition-related indices together with broader systemic inflammatory markers such as the systemic inflammation response index (SIRI) and aggregate index of systemic inflammation (AISI) in women with low AMH remains limited.

The present study was not designed to identify alternatives to established ovarian reserve markers such as AMH or AFC or to predict reproductive outcomes. Rather, we hypothesized that women with low AMH may exhibit a distinct systemic inflammatory and immunonutritional phenotype. Accordingly, we investigated whether routinely derived inflammatory and immunonutritional indices differed according to AMH status, correlated with continuous AMH concentrations, and remained associated with low AMH after adjustment for age and BMI.

Therefore, the present study aimed to evaluate systemic inflammatory and nutrition-related indices in women with low AMH levels compared with women with higher AMH levels. Specifically, we investigated NLR, PLR, SII, SIRI, AISI, PNI, and HALP; examined their correlations with continuous serum AMH concentrations; and assessed their associations with low AMH after adjustment for age and BMI.

Materials and Methods

Ethical considerations

The requirement for written informed consent was waived by the Ethics Committee because of the retrospective study design and use of previously collected clinical and laboratory data. All data were analyzed in anonymized form." Please give the complete ethics approval number as E-22686390-200-102527 (29 June 2026).

Study design and participants

This retrospective cross-sectional study included 150 women of reproductive age who were evaluated at the Department of Obstetrics and Gynecology of Istanbul Atlas University Hospital between 15 June 2025 and 15 June 2026. Clinical, anthropometric, hormonal, biochemical, and hematological data were obtained from the participants' medical records and laboratory results.

The study population was divided into two groups according to serum AMH concentrations: the low AMH group (n=75) and the high AMH group (n=75). The AMH threshold used for group allocation was defined before the statistical analyses in accordance with the study protocol. Age, body mass index (BMI), reproductive history, menstrual characteristics, antral follicle count (AFC), hormonal measurements, metabolic parameters, serum proteins, and complete blood count variables were evaluated.

Women of reproductive age who underwent serum AMH assessment and had complete clinical and laboratory data required for the study were considered eligible. Women with conditions or treatments potentially affecting serum AMH

concentrations or systemic inflammatory, metabolic, hematological, or immunonutritional parameters were excluded. Specifically, women with polycystic ovary syndrome (PCOS), endometriosis or endometrioma, previous ovarian surgery, premature ovarian insufficiency, autoimmune disease, acute or chronic infection, obesity, smoking, metabolic disease, systemic inflammatory disease, or current use of hormonal contraceptives were excluded. These eligibility criteria were applied before group allocation.

Eligible women were evaluated consecutively during the study period. After application of the predefined inclusion and exclusion criteria, the first 75 eligible women with complete clinical and laboratory data in the low AMH group and the first 75 eligible women with complete clinical and laboratory data in the high AMH group were included in the final analysis. The same selection procedure was applied to both groups.

Participants were classified into low AMH and high AMH groups using a prespecified serum AMH threshold of 1.1 ng/mL. This threshold was selected based on previously published criteria in which AMH values within approximately 0.5–1.1 ng/mL have been considered indicative of an abnormal ovarian reserve test [10]. In the present study, however, the threshold was used solely to define AMH-based study groups and was not intended to establish a clinical diagnosis of diminished ovarian reserve.

Clinical and ultrasonographic assessment

Age, BMI, gravida, parity, age at menarche, and menstrual cycle length were recorded for each participant. BMI was calculated by dividing body weight in kilograms by the square of height in meters.

AFC was assessed using transvaginal ultrasonography during the early follicular phase of the menstrual cycle. AFC was defined as the total number of follicles measuring 2–10 mm in diameter in both ovaries. To reduce interobserver variability, ultrasonographic examinations were performed by an experienced gynecologist using a standardized assessment protocol.

Blood sampling and laboratory measurements

Venous blood samples were collected in the morning after an overnight fast. Serum samples were separated by centrifugation and analyzed according to the routine procedures of the clinical biochemistry laboratory.

Complete blood count (CBC) results were generated using an automated hematology analyzer (Sysmex XN-1000; Sysmex, Norderstedt, Germany).

Routine biochemical measurements were performed in the hospital's central laboratory using enzymatic assay techniques on an automated COBAS 8000 clinical chemistry system (Roche Diagnostics, Tokyo, Japan).

Routine hormonal parameters, including AMH, FSH, LH, estradiol (E2), prolactin, and insulin levels, were measured using a Cobas e601 immunoassay analyzer (Roche Diagnostics, Basel, Switzerland).

Insulin resistance was estimated using the homeostasis model assessment of insulin resistance:

$\text{HOMA-IR} = [\text{fasting insulin } (\mu\text{IU/mL}) \times \text{fasting glucose (mg/dL)}] / 405$

All measurements were performed in accordance with the manufacturers' instructions and the laboratory's internal quality-control procedures.

Calculation of inflammatory and nutritional indices

Inflammatory and nutritional indices were derived from complete blood count and serum albumin measurements using the following equations:

$\text{NLR} = \text{neutrophil count} / \text{lymphocyte count}$

$\text{PLR} = \text{platelet count} / \text{lymphocyte count}$

$\text{SII} = (\text{platelet count} \times \text{neutrophil count}) / \text{lymphocyte count}$

$\text{SIRI} = (\text{neutrophil count} \times \text{monocyte count}) / \text{lymphocyte count}$

$\text{AISI} = (\text{platelet count} \times \text{neutrophil count} \times \text{monocyte count}) / \text{lymphocyte count}$

$\text{PNI} = [10 \times \text{serum albumin (g/dL)}] + [0.005 \times \text{total lymphocyte count } (/mm^3)]$

$\text{HALP} = (\text{hemoglobin} \times \text{albumin} \times \text{lymphocyte count}) / \text{platelet count}$

Unit conversion was applied where necessary to ensure consistency between absolute blood cell counts and the calculated index values. All derived indices were independently recalculated and compared with the values recorded in the dataset. No discrepancy exceeding an absolute tolerance of 1×10^{-6} was detected.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Continuous variables were examined separately in each AMH group using the Shapiro-Wilk test. Variables satisfying the normality assumption were summarized as mean $\pm$ standard deviation, whereas non-normally distributed variables were expressed as median and interquartile range.

Normally distributed variables were compared between the low AMH and high AMH groups using Welch's independent-samples t-test, which does not require equality of variances. Variables with non-normal distributions were evaluated using the Mann-Whitney U test. All comparisons were two-sided.

Cohen's d was calculated as the effect-size measure for parametric comparisons. Rank-biserial correlation was used for comparisons performed with the Mann-Whitney U test. To limit the false-positive rate resulting from multiple comparisons, p-values obtained from the group analyses were additionally adjusted using the Benjamini-Hochberg false discovery rate procedure. Both unadjusted p-values and adjusted q-values were reported.

Associations between serum AMH concentration and AISI, SIRI, PLR, NLR, SII, PNI, and HALP were examined using Spearman's rank correlation analysis.

Binary logistic regression analyses were performed with low AMH status coded as 1 and high AMH status coded as 0. Each inflammatory or nutritional index was evaluated in a separate model to avoid multicollinearity arising from their shared hematological components. All models were adjusted for age and BMI. Continuous predictors were standardized, and adjusted odds ratios were reported for each one-standard-deviation increase, together with their 95% confidence intervals. To further account for the potential confounding effects of age and BMI, partial Spearman correlation analyses were performed between continuous serum AMH concentrations and each inflammatory or immunonutritional index while controlling for age and BMI. Two-sided *p*-values were reported, and statistical significance was defined as *p*<0.05.

Results

A total of 150 women were included in the study, with 75 participants in the low AMH group and 75 in the high AMH group. Women in the low AMH group were significantly older than those in the high AMH group (33.0 [29.5–36.5] vs. 27.0 [24.0–31.0] years; *p*<0.001, BH-*q*<0.001). In contrast, there were no significant differences between the groups in body mass index (BMI), gravidity, parity, age at menarche, or cycle length (all *q*>0.05).

As expected, serum AMH concentration and AFC were markedly lower in the low AMH group (both *p*<0.001, BH-*q*<0.001). Women with low AMH had significantly higher FSH and LH levels and lower estradiol (E2) levels than those with high AMH (all *p*<0.001, BH-*q*<0.001). No significant between-group differences were observed in TSH or prolactin levels. Similarly, glucose, insulin, and HOMA-IR did not differ significantly between the groups.

Regarding biochemical and hematological parameters, total protein and albumin concentrations were significantly lower in the low AMH group (both *p*<0.001, BH-*q*<0.001). Neutrophil counts were significantly higher (*p*=0.001, BH-*q*=0.003), whereas lymphocyte counts were significantly lower (*p*=0.002, BH-*q*=0.003) in women with low AMH. No significant differences were detected in hemoglobin, hematocrit, monocyte count, or platelet count.

Systemic inflammatory and nutrition-related indices

All evaluated systemic inflammatory indices were significantly higher in the low AMH group than in the high AMH group. AISI was 299 [238–354] versus 227 [181–298] (*p*<0.001), SIRI was 1.08 [0.90–1.33] versus 0.89 [0.68–1.08] (*p*<0.001), PLR was 141.41±29.93 versus 127.17±27.18 (*p*=0.003), NLR was 2.26±0.48 versus 1.92±0.53 (*p*<0.001), and SII was 596 [500–692] versus 465 [398–602] (*p*<0.001). All of these differences remained statistically significant after Benjamini-Hochberg correction for multiple comparisons.

In contrast, nutrition-related indices were significantly lower in women with low AMH. PNI was 51.36±2.23 in the low AMH group compared with 53.91±2.38 in the high AMH group

(*p*<0.001), while HALP was 39.49 [31.78–44.81] versus 45.35 [38.42–50.79], respectively (*p*<0.001). Both differences remained significant after false-discovery-rate correction (BH-*q*<0.001) (Table 1).

Correlations between AMH concentration and inflammatory/nutritional indices

Spearman correlation analysis demonstrated significant inverse correlations between continuous serum AMH concentration and AISI (*ρ*=−0.282, *p*<0.001), SIRI (*ρ*=−0.266, *p*=0.001), PLR (*ρ*=−0.257, *p*=0.001), NLR (*ρ*=−0.314, *p*<0.001), and SII (*ρ*=−0.324, *p*<0.001). In contrast, AMH concentration was positively correlated with PNI (*ρ*=0.466, *p*<0.001) and HALP (*ρ*=0.366, *p*<0.001). Among the evaluated indices, the strongest correlation with AMH was observed for PNI (Table 2).

Age- and BMI-adjusted associations with low AMH

In separate binary logistic regression models adjusted for age and BMI, higher levels of all systemic inflammatory indices were significantly associated with greater odds of low AMH. For each 1-standard-deviation increase, the adjusted odds ratio (OR) was 1.78 for AISI (95% CI: 1.18–2.69; *p*=0.006), 1.90 for SIRI (95% CI: 1.26–2.89; *p*=0.002), 1.68 for PLR (95% CI: 1.14–2.49; *p*=0.009), 2.03 for NLR (95% CI: 1.34–3.06; *p*<0.001), and 1.87 for SII (95% CI: 1.25–2.81; *p*=0.002).

Conversely, higher PNI and HALP values were independently associated with lower odds of low AMH. The adjusted OR was 0.31 for PNI (95% CI: 0.19–0.51; *p*<0.001) and 0.52 for HALP (95% CI: 0.34–0.79; *p*=0.002). Among the evaluated models, the PNI model had the lowest Akaike information criterion value (AIC=159.3) (Table 3).

After adjustment for age and BMI, the associations between continuous AMH concentrations and all investigated indices remained statistically significant. AMH was inversely associated with AISI (partial *ρ*=−0.261, *p*=0.001), SIRI (partial *ρ*=−0.254, *p*=0.002), PLR (partial *ρ*=−0.234, *p*=0.004), NLR (partial *ρ*=−0.273, *p*<0.001), and SII (partial *ρ*=−0.282, *p*<0.001). In contrast, AMH remained positively associated with PNI (partial *ρ*=0.385, *p*<0.001) and HALP (partial *ρ*=0.321, *p*<0.001). These findings indicate that the observed relationships were attenuated only modestly after adjustment for age and BMI.

Discussion

The most important finding of the present study is that women with low AMH exhibited a consistent systemic inflammatory and immunonutritional profile characterized by higher AISI, SIRI, PLR, NLR, and SII and lower PNI and HALP levels. Importantly, although women in the low AMH group were significantly older, each of these indices remained significantly associated with low AMH when evaluated in separate models adjusted for age and BMI. Among the inflammatory indices, NLR showed the strongest positive association with low AMH, with a 1-SD increase corresponding to approximately twofold higher odds of low AMH (adjusted OR 2.03). In con-

Table 1. Participant characteristics according to AMH group

Variable	Low AMH (n=75)	High AMH (n=75)	p	BH q	Effect size
Age (years)	33.00 [29.50–36.50]	27.00 [24.00–31.00]	<0.001	<0.001	0.51
AMH (ng/mL)	0.55 [0.17–0.83]	3.39 [2.58–4.79]	<0.001	<0.001	-1.00
BMI (kg/m ²)	24.86 [22.72–28.09]	25.54 [23.19–27.57]	0.774	0.799	0.03
Gravida (n)	2.00 [0.00–3.00]	2.00 [1.00–3.00]	0.131	0.203	-0.14
Parity (n)	1.00 [0.00–1.00]	1.00 [0.00–2.00]	0.203	0.286	-0.11
Age at menarche (years)	12.00 [12.00–14.00]	13.00 [11.00–14.00]	0.991	0.991	0.00
Cycle length (days)	30.00 [28.00–31.00]	29.00 [28.00–31.00]	0.516	0.593	0.06
FSH (mIU/mL)	13.46 [10.86–15.79]	6.49 [5.37–7.88]	<0.001	<0.001	1.00
LH (mIU/mL)	8.29 [5.84–9.63]	5.74 [4.33–7.53]	<0.001	<0.001	0.50
E2 (pg/mL)	70 [52–81]	91 [69–104]	<0.001	<0.001	-0.50
AFC (n)	3.00 [2.00–4.00]	13.00 [11.00–16.00]	<0.001	<0.001	-1.00
TSH (mIU/L)	2.01 [1.16–2.67]	2.10 [1.56–2.72]	0.346	0.444	-0.09
Prolactin (ng/mL)	15.06 [9.78–19.55]	14.23 [10.36–18.77]	0.666	0.711	0.04
Glucose (mg/dL)	91 [87–96]	90 [86–95]	0.492	0.586	0.07
Insulin (μIU/mL)	10.06 [7.54–12.05]	10.53 [8.38–12.52]	0.090	0.154	-0.16
HOMA-IR	2.22 [1.67–2.64]	2.38 [1.85–2.84]	0.096	0.157	-0.16
Total protein (g/dL)	6.91±0.21	7.07±0.26	<0.001	<0.001	-0.69
Albumin (g/dL)	4.16±0.15	4.33±0.17	<0.001	<0.001	-1.08
Hb (g/dL)	12.80 [12.34–13.09]	12.88 [12.27–13.17]	0.358	0.444	-0.09
Hct (%)	38.40±1.86	38.55±1.83	0.615	0.681	-0.08
Monocyte (10 ³ /μL)	0.50 [0.46–0.54]	0.49 [0.41–0.54]	0.176	0.259	0.13
Platelet (10 ³ /μL)	269.04±33.70	262.87±31.03	0.245	0.330	0.19
Neutrophil (10 ³ /μL)	4.31±0.65	3.94±0.73	0.001	0.003	0.54
Lymphocyte (10 ³ /μL)	1.95 [1.71–2.15]	2.07 [1.87–2.39]	0.002	0.003	-0.30
AISI (index)	299 [238–354]	227 [181–298]	<0.001	<0.001	0.34
SIRI	1.08 [0.90–1.33]	0.89 [0.68–1.08]	<0.001	<0.001	0.35
PLR	141.41±29.93	127.17±27.18	0.003	0.005	0.50
NLR	2.26±0.48	1.92±0.53	<0.001	<0.001	0.67
SII	596 [500–692]	465 [398–602]	<0.001	<0.001	0.37
PNI	51.36±2.23	53.91±2.38	<0.001	<0.001	-1.11
HALP	39.49 [31.78–44.81]	45.35 [38.42–50.79]	<0.001	<0.001	-0.39

Data are presented as mean±standard deviation (SD) or median [interquartile range (IQR)], according to the distribution. Between-group comparisons were performed using Welch's independent-samples t-test or the Mann-Whitney U test, as appropriate. BH q denotes the Benjamini-Hochberg false-discovery-rate-adjusted p value. Effect sizes are reported as Cohen's d for parametric comparisons and rank-biserial r for non-parametric comparisons; positive values indicate higher levels in the low AMH group. Composite indices are expressed in calculated index units. AFC: Antral follicle count; AISI: Aggregate index of systemic inflammation; AMH: Anti-Müllerian hormone; BMI: Body mass index; E2: Estradiol; FSH: Follicle-stimulating hormone; HALP: Hemoglobin-albumin-lymphocyte-platelet index; HOMA-IR: Homeostasis model assessment of insulin resistance; LH: Luteinizing hormone; NLR: Neutrophil-to-lymphocyte ratio; PLR: Platelet-to-lymphocyte ratio; PNI: Prognostic nutritional index; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index; TSH: Thyroid-stimulating hormone.

trast, PNI showed the strongest inverse association (adjusted OR 0.31). Furthermore, continuous AMH concentrations were inversely correlated with all inflammatory indices and positively correlated with PNI and HALP. Collectively, these findings suggest that low AMH is accompanied by a broader pattern of increased systemic inflammatory activity and an altered immunonutritional profile rather than by changes in a single inflammatory marker.

Our findings regarding systemic inflammatory indices are supported by emerging evidence linking peripheral inflammatory activity to ovarian reserve. In a large retrospective cohort of 19,282 women undergoing assisted reproductive treatment, Yao et al. [7] reported that higher SII, PLR, and NLR were associated with lower AMH levels, lower AFC, and fewer retrieved

oocytes. These associations persisted in adjusted analyses, and dose-response relationships were observed across inflammatory-marker categories. Notably, the reported relationship was not significantly modified by age or BMI, lending support to the possibility that systemic inflammatory status may provide information beyond these major determinants of ovarian reserve. The concordance between that large cohort and the present findings is particularly relevant because our study similarly demonstrated higher SII, PLR, and NLR in women with low AMH and significant inverse correlations of these indices with continuous AMH levels.

The association observed for NLR is also consistent with earlier findings in women with impaired ovarian function. Ilhan et al. [5] reported significantly higher NLR values in women with

Table 2. Spearman and age- and BMI-adjusted partial correlations between continuous AMH concentrations and inflammatory/immunonutritional indices

Index	Spearman ρ	p	Age- and BMI-adjusted partial ρ	p
AISI	-0.282	<0.001	-0.261	0.001
SIRI	-0.266	0.001	-0.254	0.002
PLR	-0.257	0.001	-0.234	0.004
NLR	-0.314	<0.001	-0.273	<0.001
SII	-0.324	<0.001	-0.282	<0.001
PNI	0.466	<0.001	0.385	<0.001
HALP	0.366	<0.001	0.321	<0.001

Values are Spearman's rank correlation coefficients (ρ) and age- and BMI-adjusted partial Spearman correlation coefficients with two-sided p values. Partial correlations were calculated while controlling for age and BMI. Negative coefficients indicate inverse associations with AMH, whereas positive coefficients indicate direct associations. AISI: Aggregate index of systemic inflammation; AMH: Anti-Müllerian hormone; HALP: Hemoglobin-albumin-lymphocyte-platelet index; NLR: Neutrophil-to-lymphocyte ratio; PLR: Platelet-to-lymphocyte ratio; PNI: Prognostic nutritional index; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index; BMI: Body mass index.

premature ovarian insufficiency and found a negative correlation between NLR and AMH ($r=-0.27$). NLR demonstrated an AUC of 0.66 for premature ovarian insufficiency in that study [5]. In the present study, NLR was similarly inversely correlated with AMH ($\rho=-0.314$) and remained significantly associated with low AMH after adjustment for age and BMI. Although premature ovarian insufficiency and low AMH are clinically distinct entities and should not be considered interchangeable, the consistency in the direction of these associations supports a potential relationship between altered systemic immune-cell balance and low AMH.

Not all studies, however, have demonstrated an association between systemic inflammatory markers and ovarian reserve. Shoarishoar et al. [6] evaluated 174 women classified according to AMH concentrations above or below 1.1 ng/mL and found no significant associations between diminished ovarian reserve and NLR, PLR, or the red cell distribution width-to-platelet ratio; age and FSH were the principal variables associated with diminished ovarian reserve. Similarly, Long et al. [11], in an analysis of 703 reproductive-aged women, found no significant association between circulating C-reactive protein and AMH or the odds of diminished ovarian reserve after adjustment. Previous studies evaluating routine inflammatory indices in premature ovarian insufficiency (POI) have yielded heterogeneous findings [12–15]. Sanverdi et al. [12] reported higher NLR in women with POI, while Baki Erin demonstrated higher NLR, PLR, and SII and inverse correlations of these indices with AMH and antral follicle count. In contrast, Demir et al. [13] found no significant differences in NLR or PLR between women with POI and controls, and Yldrm et al. [15] unexpectedly reported lower NLR in POI. These discrepant findings may reflect differences in study populations, definitions of ovarian dysfunction, and inflammatory markers evaluated. Our results, showing higher NLR, PLR, SII, SIRI, and AISI and inverse correlations of

Table 3. Age- and BMI-adjusted associations with low AMH

Predictor	Adjusted OR (per 1 SD)	95% CI	p	AIC
AISI	1.78	1.18–2.69	0.006	179.1
SIRI	1.90	1.26–2.89	0.002	177.0
PLR	1.68	1.14–2.49	0.009	180.3
NLR	2.03	1.34–3.06	<0.001	174.7
SII	1.87	1.25–2.81	0.002	177.4
PNI	0.31	0.19–0.51	<0.001	159.3
HALP	0.52	0.34–0.79	0.002	177.2

Each index was evaluated in a separate binary logistic regression model with low AMH coded as 1 and High AMH coded as 0. Models were adjusted for age and BMI. Adjusted odds ratios (ORs) and 95% confidence intervals (CIs) are reported per 1-standard-deviation (SD) increase in the predictor. Indices were not entered simultaneously because they share hematological components. AMH: Anti-Müllerian hormone; AIC: Akaike information criterion; AISI: Aggregate index of systemic inflammation; BMI: Body mass index; CI: Confidence interval; HALP: Hemoglobin-albumin-lymphocyte-platelet index; NLR: Neutrophil-to-lymphocyte ratio; OR: Odds ratio; PLR: Platelet-to-lymphocyte ratio; PNI: Prognostic nutritional index; SII: Systemic immune-inflammation index; SIRI: Systemic inflammation response index.

these indices with AMH, add to the evidence supporting an association between reduced ovarian reserve and a systemic inflammatory profile. These differences may reflect heterogeneity in study populations, definitions of reduced ovarian reserve, inflammatory markers examined, timing of blood collection, comorbidities, and statistical adjustment. They may also indicate that different inflammatory biomarkers capture distinct components of immune activation. Unlike a single acute-phase protein such as CRP, cell-derived composite indices integrate neutrophil, lymphocyte, monocyte, and platelet compartments and therefore may characterize systemic inflammatory balance differently.

Similar findings have recently been reported in women with endometriosis. Ulger et al. [16] evaluated 91 women with surgically confirmed endometriomas and demonstrated significant inverse correlations between serum AMH concentrations and PLR, SII, and the monocyte-to-lymphocyte ratio. Moreover, age and PLR remained independently associated with AMH levels in multivariable analysis. These findings are consistent with our results, in which AMH was inversely correlated with PLR, SII, NLR, SIRI, and AISI, while higher inflammatory indices were associated with greater odds of low AMH after adjustment for age and BMI. The concordant findings support an association between systemic inflammatory status and ovarian reserve. However, because the study by Ulger et al. [16] was restricted to women with endometriomas, in whom local and systemic inflammation is an established component of the underlying disease, its findings cannot be directly generalized to women with low AMH without endometriosis. Similarly, Cui et al. [17] reported lower serum AMH concentrations in women with bilateral tubal occlusion used as a clinical model of chronic pelvic inflammation, with the difference remaining significant after BMI adjustment. Further clinical evidence supporting the relationship between inflammation and ovarian reserve has recently been provid-

ed by Griffiths et al. [18]. In women undergoing diagnostic laparoscopy, serum AMH concentrations were significantly lower in those with superficial peritoneal endometriosis, both with and without concomitant ovarian endometriosis, compared with women without endometriosis, even after accounting for age. Importantly, reduced AMH in women with superficial peritoneal endometriosis was accompanied by an elevated inflammatory profile, including increased serum IL-17 and TNF- α and increased peritoneal fluid IL-23 [18]. These observations are consistent with our findings demonstrating higher systemic inflammatory indices in women with low AMH and inverse correlations of AMH with NLR, PLR, SII, SIRI, and AISI. Although the inflammatory milieu associated with endometriosis represents a specific disease context and therefore cannot be directly extrapolated to women with low AMH in general, the findings strengthen the concept that reduced ovarian reserve may coexist with enhanced systemic and local inflammatory activity. These findings are consistent with the concept that both localized and systemic inflammatory environments may accompany impaired ovarian reserve. Nevertheless, studies in endometriosis or chronic pelvic inflammatory conditions cannot be directly extrapolated to otherwise unselected women with low AMH because the underlying inflammatory burden and mechanisms may differ substantially.

However, not all studies have demonstrated an independent association between peripheral inflammatory markers and AMH. In a recent retrospective cross-sectional study of 819 reproductive-age women, Karalök et al. [19] found that neither NLR nor CRP was independently associated with serum AMH after adjustment for chronological age. In particular, NLR showed no significant age-adjusted association with AMH, suggesting that the relationship observed in unselected reproductive-age populations may be largely influenced by age or may not be adequately captured by routine peripheral inflammatory markers [18]. This contrasts with our findings, in which NLR was inversely correlated with continuous AMH concentrations and remained significantly associated with low AMH after adjustment for both age and BMI. Moreover, our study demonstrated consistent associations across several complementary inflammatory indices, including AISI, SIRI, PLR, and SII. Differences in study design, participant characteristics, definition of ovarian reserve, statistical modeling, and the spectrum of inflammatory indices evaluated may partly account for these discrepant findings. Importantly, the absence of an association with isolated NLR or CRP does not necessarily exclude a relationship between ovarian reserve and broader systemic inflammatory or immunonutritional profiles.

A biological relationship between inflammation and ovarian aging is plausible, although the present findings cannot establish causality. Ovarian aging is a multifactorial process, and previous studies have described age-related alterations in inflammatory and immune pathways within the ovarian microenvironment [3, 4, 20]. Zeng et al. [3] highlighted chronic inflammation as a potentially important component of ovarian

aging, while Seyhanli et al. [20] described age-related increases in ovarian inflammatory mediators and immune-cell accumulation. These observations provide biological context for the inverse associations between AMH and AISI, SIRI, NLR, PLR, and SII observed in the present study. However, our cross-sectional data cannot determine whether systemic inflammation contributes to declining ovarian reserve, develops in parallel with reproductive aging, or reflects shared age-related biological processes. Accordingly, these findings should be interpreted primarily as an association between lower AMH and a more pronounced systemic inflammatory profile rather than evidence of a causal mechanism.

An important and relatively novel aspect of the present study is the simultaneous evaluation of nutrition-related composite indices. Compared with conventional inflammatory markers such as NLR, PLR, and SII, PNI and HALP have been considerably less studied in relation to ovarian reserve. PNI combines serum albumin and lymphocyte count and therefore reflects both nutritional and immune-inflammatory components, whereas HALP integrates hemoglobin, albumin, lymphocyte, and platelet values [8, 9]. Both indices have increasingly been used as composite immunonutritional markers in different clinical settings, including reproductive and obstetric populations [20]. In the present study, both PNI and HALP were significantly lower in women with low AMH and were positively correlated with continuous AMH concentrations. Moreover, both remained inversely associated with low AMH after adjustment for age and BMI, with PNI demonstrating the strongest association and lowest AIC among the evaluated models. These findings extend the concept of altered systemic status in women with low AMH from inflammation alone toward a broader immunonutritional phenotype.

The PNI finding warrants particular attention. AMH showed its strongest correlation among the investigated indices with PNI ($p=0.466$), and each 1-SD increase in PNI was associated with 69% lower odds of low AMH after adjustment for age and BMI. However, PNI should not be interpreted as a direct measure of dietary intake or clinical malnutrition. Albumin is influenced by inflammatory and metabolic processes in addition to nutritional status, and lymphocyte counts reflect immune status. Thus, in this setting, PNI may be better regarded as an integrated immunonutritional marker rather than a specific measure of nutritional deficiency. Similarly, the lower HALP values may partly reflect the lower albumin and lymphocyte levels because these variables are mathematical components of the index.

The present findings should be interpreted as demonstrating an association between low AMH and a systemic inflammatory-immunonutritional phenotype rather than as evidence that the evaluated indices improve ovarian reserve assessment beyond established measures such as AMH and AFC. Because the study groups were defined according to AMH concentration, the present data cannot establish incremental diagnostic utility. Similarly, reproductive outcomes were not evaluated, and no prognostic inference can be made.

Strengths and limitations

The present study has several strengths. A broad panel of routinely available systemic inflammatory and immunonutritional indices was evaluated within the same population, allowing their associations with AMH status and continuous AMH concentrations to be compared directly. AMH was assessed both as a categorical and continuous variable, and the analytical approach included correlation analysis, age- and BMI-adjusted partial correlation analysis using continuous AMH, and age- and BMI-adjusted logistic regression. In addition, Benjamini–Hochberg correction was applied to baseline comparisons to reduce the risk of false-positive findings due to multiple testing. Nevertheless, several limitations should be acknowledged. First, the retrospective cross-sectional design precludes determination of temporal or causal relationships between systemic inflammation, immunonutritional status, and low AMH. Second, women with low AMH were significantly older than those with high AMH. Although the associations between continuous AMH and all investigated inflammatory and immunonutritional indices remained statistically significant after adjustment for age and BMI, residual confounding related to reproductive aging cannot be completely excluded. Therefore, these findings should be interpreted as age- and BMI-adjusted associations rather than evidence that the observed systemic alterations are fully independent of reproductive aging. Third, inflammatory and nutritional indices and AMH were evaluated at a single time point; therefore, potential intra-individual changes over time could not be assessed. Fourth, the adjusted models included age and BMI, but residual confounding from other factors that may influence systemic inflammation or AMH, including lifestyle factors, unmeasured medication use, and underlying comorbidities, cannot be excluded. Furthermore, several of the evaluated indices share common hematological components and are therefore not biologically independent. Accordingly, the indices were examined in separate regression models, limiting assessment of their independent contributions when considered simultaneously. Finally, because the study population was defined according to AMH concentration and reproductive outcomes were not assessed, the present findings cannot establish incremental diagnostic or prognostic utility. Further prospective, multicenter studies are needed to confirm these associations and clarify their biological and clinical significance.

Clinical implications/future directions

The evaluated indices are inexpensive and readily derived from routine complete blood count and biochemical parameters. However, the present findings should be interpreted as reflecting the systemic inflammatory and immunonutritional context associated with low AMH rather than as evidence of diagnostic utility. In particular, PNI may warrant further investigation because it showed the strongest correlation with AMH and the strongest adjusted inverse association with low AMH. Prospective multicenter studies with longitudinal measurements are needed to confirm these associations and clarify their biological and potential clinical significance.

Conclusion

Women with low AMH exhibited a systemic profile characterized by higher inflammatory indices and lower nutrition-related indices. AMH was inversely associated with AISI, SIRI, PLR, NLR, and SII and positively associated with PNI and HALP. Importantly, these associations remained significant after adjustment for age and BMI. These findings support an association between low AMH and a distinct systemic inflammatory–immunonutritional phenotype but do not establish that these indices improve ovarian reserve assessment beyond established markers such as AMH and AFC. Prospective studies are needed to determine the biological and potential clinical significance of these associations.

Disclosures

Ethics Committee Approval: The study was approved by the Istanbul Atlas University Non-Interventional Scientific Research Clinical Research Ethics Committee (no: E-22686390-200-102527, date: 29/06/2026).

Informed Consent: The requirement for written informed consent was waived by the Ethics Committee because of the retrospective study design and use of previously collected clinical and laboratory data. All data were analyzed in anonymized form.

Conflict of Interest Statement: None declared.

Funding: The authors declare that this study received no financial support.

Use of AI for Writing Assistance: OpenAI ChatGPT was used for language editing, manuscript organization, and proofreading assistance. All AI-assisted text was critically reviewed and revised by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Authorship Contributions: Concept – O.D.A., N.F.M., H.U.; Design – O.D.A.; Supervision – O.D.A.; Data collection and/or processing – O.D.A.; Data analysis and/or interpretation – N.F.M., H.U.; Writing – O.D.A., N.F.M., H.U.; Critical review – O.D.A., N.F.M., H.U.

Peer-review: Externally peer-reviewed.

References

1. Moolhuijsen LME, Visser JA. Anti-Müllerian Hormone and Ovarian Reserve: Update on Assessing Ovarian Function. *J Clin Endocrinol Metab* 2020;105(11):3361–73. [\[CrossRef\]](#)
2. Broer SL, Broekmans FJ, Laven JS, Fauser BC. Anti-Müllerian hormone: ovarian reserve testing and its potential clinical implications. *Hum Reprod Update* 2014;20(5):688–701. [\[CrossRef\]](#)
3. Zeng Y, Wang C, Yang C, Shan X, Meng XQ, Zhang M. Unveiling the role of chronic inflammation in ovarian aging: insights into mechanisms and clinical implications. *Hum Reprod* 2024;39(8):1599–607. [\[CrossRef\]](#)
4. Yan F, Zhao Q, Li Y, Zheng Z, Kong X, Shu C, et al. The role of oxidative stress in ovarian aging: a review. *J Ovarian Res* 2022;15(1):100. [\[CrossRef\]](#)

5. İlhan G, Atmaca FFV, Altan E, Zebitay AG, Sozen H, Akyol H, et al. Evaluation of neutrophil-lymphocyte ratio, platelet-lymphocyte ratio and red blood cell distribution width-platelet ratio for diagnosis of premature ovarian insufficiency. *J Family Reprod Health* 2016;10(4):211–6.
6. Shoarishoar SS, KaboodMehri R, Fakor F, Rafiei Sorouri Z, Mansour-Ghanaei M, Faraji Darkhaneh Ret al. Assessment of decreased ovarian reserve and systemic inflammatory markers. *Cell Mol Biol (Noisy-le-grand)* 2024;70(11):144–9. [\[CrossRef\]](#)
7. Yao QZ, Liu RL, Zeng YY, Lv XY, Geng LH, Hu YL, et al. Investigating the association between systemic immune-inflammation index and female fertility: a retrospective cohort study. *J Ovarian Res* 2026;19(1):47. [\[CrossRef\]](#)
8. Yan L, Nakamura T, Casadei-Gardini A, Bruixola G, Huang YL, Hu ZD. Long-term and short-term prognostic value of the prognostic nutritional index in cancer: a narrative review. *Ann Transl Med* 2021;9(21):1630. [\[CrossRef\]](#)
9. Farag CM, Antar R, Akosman S, Ng M, Whalen MJ. What is hemoglobin, albumin, lymphocyte, platelet (HALP) score? A comprehensive literature review of HALP's prognostic ability in different cancer types. *Oncotarget* 2023;14:153–72. [\[CrossRef\]](#)
10. Ferraretti AP, La Marca A, Fauser BC, Tarlatzis B, Nargund G, Gianaroli L. ESHRE consensus on the definition of 'poor response' to ovarian stimulation for *in vitro* fertilization: the Bologna criteria. *Hum Reprod* 2011;26(7):1616–24. [\[CrossRef\]](#)
11. Long A, Steiner AZ, Thompson AL, Jahnke HR, Harris BS, Jukic AM. Inflammation and ovarian function in reproductive-aged women. *Am J Hum Biol* 2025;37(1):e24196. [\[CrossRef\]](#)
12. Sanverdi I, Kilicci C, Cogendez E, Abide Yayla C, Ozkaya E. Utility of complete blood count parameters to detect premature ovarian insufficiency in cases with oligomenorrhea/amenorrhea. *J Clin Lab Anal* 2018;32(5):e22372. [\[CrossRef\]](#)
13. Demir B, Sarıdaş Demir S, Özkan Karacaer K, Paşa S, Silan F. Evaluation inflammatory markers of hemogram parameters in primary ovarian insufficiency. *Turk J Obstet Gynecol* 2020;17(1):9–14. [\[CrossRef\]](#)
14. Baki Erin K. May the Systemic Immune-Inflammation Index be an Indicator of Premature Ovarian Insufficiency? *Middle Black Sea J Heal Sci* 2023;9(1):125–34. [\[CrossRef\]](#)
15. Yldrm G, Tokmak A, Kokanal MK, Sarkaya E, Züngün C, İnal HA, et al. Association between some inflammatory markers and primary ovarian insufficiency. *Menopause* 2015;22(9):1000–5. [\[CrossRef\]](#)
16. Ulger G, Yildiz H, Akay K, Yildizbakan A, Aytan P, Yazici FG, et al. Is systemic inflammation associated with reduced ovarian reserve in endometriosis? An investigation of anti-Müllerian hormone and inflammatory markers. *Int J Gynaecol Obstet* 2026;173(1):296–301. [\[CrossRef\]](#)
17. Cui L, Sheng Y, Sun M, Hu J, Qin Y, Chen ZJ. Chronic pelvic inflammation diminished ovarian reserve as indicated by serum anti müllerian hormone. *PLoS One* 2016;11(6):e0156130. [\[CrossRef\]](#)
18. Griffiths MJ, Brown ME, Gibson DA, Collins F, Dunlop CE, Saunders PTK, et al. Anti-Müllerian hormone decreased in women with superficial peritoneal endometriosis and associated with an elevated inflammatory profile. *Reprod Fertil* 2026;7(2):RAF250127. [\[CrossRef\]](#)
19. Karalök MH, Dünder B, Parmaksız A, Elgun T, Gündogan GI, Gök Yurttaş A. Age-Adjusted associations between routine systemic inflammatory markers and anti-müllerian hormone in reproductive-age women: a retrospective cross-sectional study. *Biomedicines* 2026;14(8):1733. [\[CrossRef\]](#)
20. Seyhanlı Z, Bayraktar B, Karabay G, Filiz AA, Bucak M, Agaoğlu RT, et al. Can maternal inflammatory and nutritional status, evaluated by the hemoglobin, albumin, lymphocyte, and platelet (HALP) score and the prognostic nutritional index (PNI) in the first trimester, predict late-onset fetal growth restriction? *BMC Pregnancy Childbirth* 2024;24(1):620. [\[CrossRef\]](#)



Research Article

Clinical and laboratory correlates of low fecal elastase-1 levels: A retrospective analysis

Neziha Hacıhasanoğlu Cakmak¹, Gözde Ulfer², Pakize Yigit³, Turkan Yigitbasi⁴, Gulbanu Canbaloglu⁵

¹Medical Laboratory Techniques, Istanbul Medipol University Vocational School of Health Services, Istanbul, Türkiye

²Department of Medical Biochemistry, Istanbul Medipol University Faculty of Medicine, Istanbul, Türkiye

³Department of Management Information Systems, Istanbul Medipol University Faculty of Business and Management Sciences, Istanbul, Türkiye

⁴Genetic Diseases Evaluation Center, Istanbul Medipol University Faculty of Medicine, Istanbul, Türkiye

⁵Department of Gastroenterology, Istanbul Medipol University Faculty of Medicine, Istanbul, Türkiye

Abstract

Objectives: Fecal elastase-1 (FE-1) is widely used as a noninvasive, stool-based test to evaluate pancreatic exocrine function. This study compared radiological and laboratory findings between patients with low and normal FE-1 levels and investigated variables independently associated with low FE-1 levels.

Methods: The records of 1,133 patients who underwent FE-1 testing were retrospectively reviewed. Patients were classified into normal (≥ 200 $\mu\text{g/g}$, $n=961$) and low (<200 $\mu\text{g/g}$, $n=172$) FE-1 groups. Group comparisons and multivariable logistic regression analyses were performed. To assess the incremental discriminative contribution of pancreatic amylase while avoiding confounding due to differences in sample size, four additional models were evaluated using the same complete-case subset ($n=396$). Areas under the receiver operating characteristic curves (AUCs) were compared using paired DeLong tests in jamovi version 2.6.2.0.

Results: Low FE-1 levels were more frequent among patients with pathological radiological findings than among those with normal findings (31.7% vs. 12.1%; $\chi^2=45.2$, $p<0.001$; $\text{OR}=3.38$, 95% CI: 2.33–4.89). Among the same 396 complete cases, Model A (age, sex, and radiology) yielded an AUC of 0.694 (95% CI: 0.619–0.769). Adding pancreatic amylase (Model B) increased the AUC to 0.727 (95% CI: 0.657–0.798), although this increase was not statistically significant ($\Delta\text{AUC}=0.0333$; DeLong $p=0.100$). Model C (age, sex, radiology, ALT, glucose, BUN, IgA, and lipase) yielded an AUC of 0.740 (95% CI: 0.672–0.807). Adding pancreatic amylase (Model D) increased the AUC to 0.766 (95% CI: 0.705–0.827), but the increase was not statistically significant ($\Delta\text{AUC}=0.0264$; DeLong $p=0.107$).

Conclusion: Low FE-1 levels were associated with radiological, metabolic, nutritional, and laboratory abnormalities. Pancreatic amylase was independently associated with low FE-1 levels in the expanded regression model; however, its incremental contribution to AUC was not statistically significant when the models were compared using data from the same patients. These findings reflect factors associated with low FE-1 levels and should not be interpreted as establishing a diagnostic model for exocrine pancreatic insufficiency (EPI).

Keywords: Biomarkers, exocrine pancreatic insufficiency, feces, logistic models, pancreatic elastase

How to cite this article: Hacıhasanoğlu Cakmak N, Ulfer G, Yigit P, Yigitbasi T, Canbaloglu G. Clinical and laboratory correlates of low fecal elastase-1 levels: A retrospective analysis. Int J Med Biochem 2026;9(4):210–216.

Exocrine pancreatic insufficiency (EPI) develops when the pancreas fails to secrete adequate amounts of digestive enzymes, resulting in symptoms such as maldigestion, mal-

absorption, chronic diarrhea, steatorrhea, weight loss, and nutritional deficiencies [1]. The reference standards for diagnosis—the secretin stimulation test and 72-hour quantitative

Address for correspondence: Neziha Hacıhasanoğlu Cakmak, MD. Medical Laboratory Techniques, Istanbul Medipol University Vocational School of Health Services, Istanbul, Türkiye

Phone: +90 537 693 85 81 **E-mail:** ncakmak@medipol.edu.tr **ORCID:** 0000-0002-6183-5015

Submitted: September 11, 2026 **Accepted:** October 09, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



fecal fat analysis—are accurate but invasive, costly, and largely impractical outside specialized centers [2]. Consequently, fecal elastase-1 (FE-1) has emerged as the primary noninvasive screening tool for EPI in routine clinical practice [3, 4].

As a pancreatic acinar enzyme that remains largely stable during intestinal transit, FE-1 serves as a reliable indirect indicator of exocrine pancreatic function. Recent pooled meta-analytic estimates indicate a sensitivity of approximately 90–94% and a specificity of approximately 69–82% for detecting EPI using the conventional threshold of 200 µg/g, with somewhat higher sensitivity in cystic fibrosis and higher specificity in chronic pancreatitis [3, 4]. Nevertheless, its diagnostic accuracy decreases in mild-to-moderate disease.

The clinical relevance of EPI extends beyond primary pancreatic disease. Population-based screening among patients with diabetes has revealed severe EPI in 10–30% and moderate EPI in 22–56% of patients with type 1 diabetes, with prevalence rates ranging from 5% to 46% among those with type 2 diabetes, highlighting the close functional relationship between the exocrine and endocrine functions of the pancreas [5].

Although the diagnostic performance of FE-1 has been extensively investigated, relatively few studies have simultaneously examined its associations with imaging findings and a broad panel of laboratory parameters—including hematologic, iron, hepatic, renal, thyroid, lipid, and pancreatic enzyme measurements—in a large, unselected patient population. The present study aims to address this gap by comparing radiological and laboratory profiles across FE-1 categories and using multivariable modeling to identify factors independently associated with abnormal FE-1 results.

Materials and Methods

Study design and population

This single-center retrospective study drew on records from patients tested for FE-1 at the Gastroenterology Outpatient Clinic of a tertiary care university hospital across a five-year window (10 March 2020 to 10 March 2025). Of those screened, 1133 satisfied the eligibility requirements; applying the conventional threshold split them into a normal-value group (≥200 µg/g, n=961, 84.8%) and a low-value group (<200 µg/g, n=172, 15.2%). Inclusion required patients to be 18 years or older, to have a recorded FE-1 result with complete clinical documentation, and to have undergone evaluation for possible EPI. Exclusion criteria comprised insufficient clinical information required for eligibility assessment, age below 18, active flares of inflammatory bowel disease (Crohn’s disease or ulcerative colitis), a recent episode of acute gastrointestinal infection, or ongoing chemotherapy or radiotherapy for cancer. The patient flow and the samples used in the original regression models are summarized in Figure 1.

Ethical considerations

Ethical approval for this work was granted by the Institutional Non-Interventional Clinical Research Ethics Committee (Decision No: 485, Date: 26.03.2026; Reference No: E-10840098-50.04-2458), and all procedures adhered to the principles set out in the Declaration of Helsinki.

Study variables

The data collected included patient demographics (age, sex) and radiological status (pathological finding present or absent).

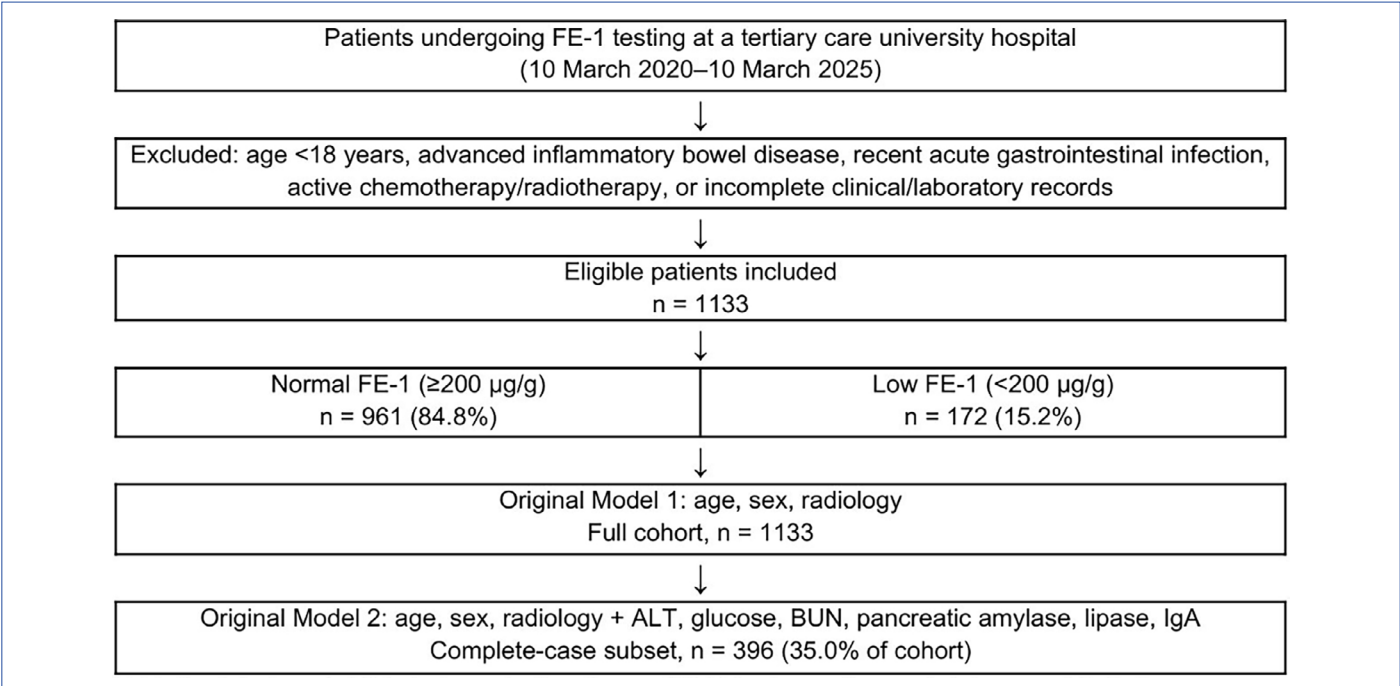


Figure 1. Patient flow diagram showing cohort selection, fecal elastase-1 grouping, and the original regression-model samples.

FE-1: Fecal elastase-1; ALT: Alanine aminotransferase; BUN: Blood urea nitrogen; IgA: Immunoglobulin A.

The laboratory panel was extensive. Hematologic parameters covered leukocyte, erythrocyte, hemoglobin, hematocrit, and erythrocyte-index values along with platelet, neutrophil, lymphocyte, monocyte, and eosinophil counts. Iron-related measures included serum iron, ferritin, unsaturated iron-binding capacity (UIBC), folic acid, vitamin B12, and vitamin D. Immunological testing captured immunoglobulin A (IgA) and anti-tissue transglutaminase IgA. Hepatic parameters spanned alanine aminotransferase (ALT, EC 2.6.1.2), aspartate aminotransferase (AST, EC 2.6.1.1), gamma-glutamyltransferase (GGT, EC 2.3.2.2), alkaline phosphatase, total and direct bilirubin, albumin, and total protein, while renal assessment relied on creatinine and blood urea nitrogen (BUN). Electrolyte values (calcium, sodium, potassium) and inflammatory indices (C-reactive protein, erythrocyte sedimentation rate) were also recorded, as were thyroid parameters (thyroid-stimulating hormone [TSH], free triiodothyronine, free thyroxine) and the lipid panel (low-density lipoprotein, high-density lipoprotein, triglycerides, total cholesterol). Pancreatic enzyme activity was assessed via pancreatic amylase (EC 3.2.1.1) and lipase (EC 3.1.1.3), and tumor markers included CA 19-9, carcinoembryonic antigen, and alpha-fetoprotein. Fecal elastase-1 (EC 3.4.21.36) was quantified using the ElaFast rapid immunochromatographic assay (Eurospital S.p.A., Trieste, Italy).

Laboratory analyses

Stool samples underwent EasyCal extraction before FE-1 quantification with a rapid immunochromatographic assay kit (ElaFast, Eurospital S.p.A., Trieste, Italy), following the manufacturer's protocol; results were classified as severe (<100 µg/g), mild-to-moderate (100–200 µg/g), or normal (>200 µg/g). A Roche cobas c 303/503/703 platform (Roche Diagnostics GmbH, Mannheim, Germany) was used for albumin, ALT and AST (International Federation of Clinical Chemistry [IFCC] methodology), C-reactive protein (Tina-quant C-Reactive Protein IV, immunoturbidimetric technique), iron (Iron Gen.2 assay) and UIBC (Unsaturated Iron-Binding Capacity assay), and IgA (Tina-quant IgA Gen.2 assay). Vitamin B12 levels were determined via electrochemiluminescence immunoassay (Elecsys Vitamin B12 II, cobas e 402/801, Roche Diagnostics GmbH, Mannheim, Germany), and anti-tissue transglutaminase IgA was quantified by enzyme-linked immunosorbent assay on the Alegria® system (ORGENTEC Diagnostika GmbH, Mainz, Germany).

Statistical analysis

Normality of continuous data was evaluated via Shapiro-Wilk testing, after which group comparisons used an independent t-test (normal data) or Mann-Whitney U test (non-normal data); chi-square testing handled categorical comparisons. For univariable laboratory comparisons, all available observations for each parameter were used; the variable-specific sample sizes for the normal and low FE-1 groups are reported in Table 1. The outcome in all logistic regression models was low FE-1 (<200 µg/g). Original Model 1 included age, sex, and radiological status and was fitted in the full cohort (n=1133). Original Model 2 included age, sex, radiological status, ALT, glucose, BUN, pancreatic amylase, lipase, and IgA and was fitted using

complete-case analysis (n=396), because only these patients had data available for all variables in that model. Because the original models used different samples, their AUCs were not interpreted as a direct estimate of incremental value. A separate complete-case analysis was therefore conducted in the same 396 patients: Model A included age, sex, and radiological findings; Model B added pancreatic amylase; Model C included Model A plus ALT, glucose, BUN, IgA, and lipase; and Model D added pancreatic amylase to Model C. AUCs with 95% confidence intervals were compared using paired DeLong tests, with B versus A and D versus C assessing the incremental discriminative contribution of pancreatic amylase. Complete cases were compared with the 737 excluded patients for age, sex, radiological status, and FE-1 category using the independent t-test or chi-square test, as appropriate. SPSS Statistics version 24.0 was used for the primary analyses and jamovi version 2.6.2.0 for paired DeLong comparisons. A p-value<0.05 was considered statistically significant; the statistical significance of AUC differences was distinguished from their magnitude and potential clinical relevance. Reporting followed STROBE guidance.

Results

Descriptive findings

Of the 1133 patients in the study cohort, 55.1% (n=624) were female and 44.9% (n=509) were male. FE-1 fell within the normal range for 84.8% (n=961) and was abnormal for 15.2% (n=172). Imaging was unremarkable in 84.1% (n=953) of cases and showed a pathological finding in the remaining 15.9% (n=180) (Fig. 1).

Categorical associations

Sex distribution did not differ meaningfully across the two FE-1 categories ($\chi^2=1.65$, df=1, p=0.198). Radiological abnormality was associated with low FE-1: 31.7% of patients with pathological imaging had low FE-1, compared with 12.1% of those with normal imaging ($\chi^2=45.2$, df=1, p<0.001, OR=3.38, 95% CI 2.33–4.89) (Table 2).

Numerical comparisons

Table 1 presents all available numerical parameters with variable-specific sample sizes for each FE-1 group. The low-FE-1 group displayed lower iron alongside higher UIBC—a pattern consistent with iron deficiency—together with elevated eosinophil count, IgA, ALT, AST, GGT, glucose, and BUN, and reduced platelet count and TSH. No meaningful difference emerged for hemoglobin, leukocyte, neutrophil, or lymphocyte counts, albumin, C-reactive protein, the lipid panel, or the other renal indices (creatinine, calcium, sodium, and potassium).

Logistic regression analyses

The original Model 1, built on age, sex, and radiological status across the full cohort (n=1133), was statistically significant (omnibus $\chi^2=42.91$, df=3, p<0.001; Nagelkerke $R^2=0.065$) and showed acceptable calibration (Hosmer-Lemeshow statistic=4.05, df=8, p=0.853), although discrimination was modest (AUC=0.635, 95% CI 0.587–0.682); pathological radiology

Table 1. Comparison of numerical parameters between normal and low fecal elastase-1 (FE-1) groups

Parameter	Normal FE-1 n	Normal FE-1 Median (IQR)	Low FE-1 n	Low FE-1 Median (IQR)	p
Age (years), median*	632	37	121	39	0.029
WBC ($\times 10^3/\mu\text{L}$)	632	7.05 (2.67)	121	6.89 (2.52)	0.559
RBC ($\times 10^6/\mu\text{L}$)	632	4.79 (0.772)	121	4.9 (0.79)	0.478
Hemoglobin (g/dL)	632	13.8 (2.4)	121	14.3 (2.5)	0.224
Hematocrit (%)	632	40.7 (6.03)	121	41.7 (5.9)	0.173
MCV (fL)	632	85.1 (5.5)	121	85.1 (6.4)	0.691
MCH (pg)	632	29.2 (2.3)	121	29 (2.5)	0.954
MCHC (g/dL)	632	34.1 (1.6)	121	34 (1.9)	0.586
Platelet count ($\times 10^3/\mu\text{L}$)	632	261 (79)	121	249 (77)	0.049
RDW-CV (%)	632	12.8 (1.1)	121	12.8 (1.1)	0.995
Neutrophil count ($\times 10^3/\mu\text{L}$)	632	3.93 (1.99)	121	3.86 (2.08)	0.196
Lymphocyte count ($\times 10^3/\mu\text{L}$)	632	2.26 (0.842)	121	2.29 (0.87)	0.443
Monocyte count ($\times 10^3/\mu\text{L}$)	632	0.55 (0.25)	121	0.54 (0.22)	0.944
Eosinophil count ($\times 10^3/\mu\text{L}$)	632	0.14 (0.14)	121	0.17 (0.18)	0.009
Iron ($\mu\text{g/dL}$)	156	84 (52.5)	26	62.5 (62.8)	0.013
Ferritin (ng/mL)	159	37.6 (84.9)	28	20.8 (62.7)	0.092
UIBC ($\mu\text{g/dL}$)	159	256 (112)	26	306 (99.3)	0.006
Folic acid (ng/mL)	151	6.4 (4.1)	25	6.6 (3.8)	0.949
Vitamin B12 (pg/mL)	153	404 (256)	25	408 (181)	0.841
Vitamin D (ng/mL)	71	25.4 (17.4)	8	21.8 (9.43)	0.398
IgA (mg/dL)	530	170 (87.8)	102	187 (105)	0.049
Anti-tTG IgA (RU/mL)	494	1 (0)	96	1 (0)	0.244
ALT (U/L)	574	20 (17)	115	25 (23)	0.004
AST (U/L)	575	20 (7)	115	22 (8)	0.021
GGT (U/L)	579	15 (16)	116	19 (19.5)	0.034
ALP (U/L)	580	65 (26)	117	65 (30)	0.217
Total bilirubin (mg/dL)	430	0.45 (0.307)	106	0.46 (0.313)	0.537
Direct bilirubin (mg/dL)	581	0.23 (0.13)	115	0.25 (0.15)	0.269
Albumin (g/dL)	556	4.76 (0.38)	111	4.77 (0.33)	0.951
Total protein (g/dL)	538	7.45 (0.54)	107	7.52 (0.515)	0.131
Glucose (mg/dL)	439	93 (11)	90	96.5 (11.8)	<0.001
Creatinine (mg/dL)	569	0.76 (0.23)	113	0.81 (0.25)	0.255
BUN (mg/dL)	563	11 (4)	113	12 (4)	0.011
Calcium (mg/dL)	560	9.71 (0.513)	112	9.79 (0.49)	0.463
Sodium (mEq/L)	553	139 (3)	109	139 (2)	0.296
Potassium (mEq/L)	562	4.42 (0.417)	110	4.42 (0.38)	0.592
C-reactive protein (mg/L)	602	1.29 (2.93)	118	1.44 (2.13)	0.449
Erythrocyte sedimentation rate (mm/h)	18	6.5 (17.3)	3	3 (14)	0.390
TSH ($\mu\text{IU/mL}$)	533	1.71 (1.27)	107	1.47 (1.03)	0.024
Free T3 (pg/mL)	434	3.12 (0.57)	100	3.14 (0.543)	0.394
Free T4 (ng/dL)	434	1.21 (0.228)	101	1.19 (0.2)	0.990
LDL (mg/dL)	371	126 (45)	79	125 (46.5)	0.913
HDL (mg/dL)	334	49 (21)	72	47.5 (21)	0.434
Triglycerides (mg/dL)	381	93 (82)	79	102 (74.5)	0.688
Total cholesterol (mg/dL)	350	194 (50)	74	188 (48.5)	0.801
Pancreatic amylase (U/L)	515	28 (15.1)	108	29 (16)	0.186
Lipase (U/L)	513	33 (15.5)	110	33.3 (20.4)	0.977
CA 19-9 (U/mL)	16	15.4 (15.4)	1	10.7 (0)	—
CEA (ng/mL)	15	1.87 (2.76)	1	1.22 (0)	—
AFP (ng/mL)	8	1.98 (1.56)	1	2.8 (0)	—

P values are based on the Mann-Whitney U test. *: Median only (25th–75th percentile not available for age). Sample sizes (n) are reported separately for each parameter because laboratory data availability varied across the retrospective cohort. For CA 19-9, CEA, and AFP, between-group p values were not calculated because only one observation was available in the low FE-1 group. IQR: Interquartile range; AFP: Alpha-fetoprotein; ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; BUN: Blood urea nitrogen; CA 19-9: Carbohydrate antigen 19-9; CEA: Carcinoembryonic antigen; FE-1: Fecal elastase-1; GGT: Gamma-glutamyltransferase; HDL: High-density lipoprotein; IgA: Immunoglobulin A; LDL: Low-density lipoprotein; MCH: Mean corpuscular hemoglobin; MCHC: Mean corpuscular hemoglobin concentration; MCV: Mean corpuscular volume; RBC: Red blood cell; RDW-CV: Red cell distribution width-coefficient of variation; T3: Triiodothyronine; T4: Thyroxine; tTG: Tissue transglutaminase; TSH: Thyroid-stimulating hormone; UIBC: Unsaturated iron-binding capacity; WBC: White blood cell.

Table 2. Categorical comparisons between normal and low fecal elastase-1 groups

Variable	Normal FE-1 n (%)	Low FE-1 n (%)	χ^2 (df)	OR (95% CI)	p
Female (n=624)	537 (86.1)	87 (13.9)	1.65 (1)	1.24 (0.89–1.71)	0.198
Male (n=509)	424 (83.3)	85 (16.7)	ref.	ref.	—
Radiology normal (n=953)	838 (87.9)	115 (12.1)	45.2 (1)	3.38 (2.33–4.89)	<0.001
Radiology pathological (n=180)	123 (68.3)	57 (31.7)	ref.	ref.	—

Percentages are row percentages (proportion within each sex/radiology subgroup that fell into the normal or low FE-1 group). FE-1: Fecal elastase-1; df: Degrees of freedom; OR: Odds ratio; CI: Confidence interval; ref.: Reference category.

Table 3. Multivariable logistic regression models for factors associated with low fecal elastase-1

Variable	Model 1			Model 2		
	OR	95% CI	p	OR	95% CI	p
Radiology (pathological)	3.221	2.216–4.684	<0.001	5.056	2.773–9.222	<0.001
Age	1.013	0.998–1.028	0.079	1.001	0.972–1.031	0.938
Sex	1.238	0.888–1.726	0.207	1.351	0.745–2.449	0.322
Glucose				1.015	0.989–1.041	0.265
BUN				1.007	0.920–1.103	0.874
Pancreatic amylase				1.031	1.004–1.059	0.026
IgA				1.003	1.000–1.007	0.065
ALT				1.008	0.997–1.019	0.155
Lipase				0.984	0.965–1.005	0.129

Model 1: age, sex, radiology (n=1133). Model 2: age, sex, radiology, ALT, glucose, BUN, pancreatic amylase, lipase, IgA (n=396). Reference categories: radiology = normal (0); sex = female. ALT: Alanine aminotransferase; BUN: Blood urea nitrogen; CI: Confidence interval; IgA: Immunoglobulin A; OR: Odds ratio.

Table 4. Discrimination of same-sample models and pairwise DeLong comparisons in complete cases (n=396)

Analysis	Model/comparison	AUC / Δ AUC	SE	95% CI	z	p
Model performance	Model A	0.694	0.0384	0.619–0.769	—	<0.001
Model performance	Model B	0.727	0.0358	0.657–0.798	—	<0.001
Model performance	Model C	0.740	0.0346	0.672–0.807	—	<0.001
Model performance	Model D	0.766	0.0310	0.705–0.827	—	<0.001
DeLong comparison	Model A vs. Model B	-0.0333	—	-0.0729–0.00638	-1.644	0.100
DeLong comparison	Model A vs. Model C	-0.0454	—	-0.0919–0.00117	-1.911	0.056
DeLong comparison	Model D vs. Model A	0.0717	—	0.02203–0.1214	2.829	0.005
DeLong comparison	Model B vs. Model C	-0.0121	—	-0.0712–0.04708	-0.400	0.689
DeLong comparison	Model B vs. Model D	-0.0384	—	-0.0817–0.00478	-1.743	0.081
DeLong comparison	Model C vs. Model D	-0.0264	—	-0.0584–0.00572	-1.611	0.107

Model A: age, sex, and radiology; Model B: Model A + pancreatic amylase; Model C: Model A + ALT, glucose, BUN, IgA, and lipase; Model D: Model C + pancreatic amylase. AUC: Area under the receiver operating characteristic curve; CI: Confidence interval; SE: Standard error; Δ AUC: Difference in area under the curve.

was the only independent correlate (Table 3). Original Model 2, fitted to the complete-case subset (n=396) and including age, sex, radiological status, ALT, glucose, BUN, pancreatic amylase, lipase, and IgA, was also significant (omnibus $\chi^2=49.48$, df=9, p<0.001; Nagelkerke $R^2=0.196$), showed good calibration (Hosmer-Lemeshow statistic=4.61, df=8, p=0.798), and had an AUC of 0.761 (95% CI 0.700–0.822). Pancreatic amylase was independently associated with low FE-1 (OR=1.031, 95% CI 1.004–1.059, p=0.026). Because these original models used different samples, their AUC difference was not used to infer incremental contribution (Fig. 1).

In the same 396 complete cases, Model A (age, sex, radiology) yielded an AUC of 0.694 (95% CI 0.619–0.769). Model B, which

additionally included pancreatic amylase, yielded an AUC of 0.727 (95% CI 0.657–0.798); the AUC increase of 0.0333 was not statistically significant (DeLong p=0.100). Model C (age, sex, radiology, ALT, glucose, BUN, IgA, and lipase) yielded an AUC of 0.740 (95% CI 0.672–0.807). Model D, which added pancreatic amylase to Model C, yielded an AUC of 0.766 (95% CI 0.705–0.827); the AUC increase of 0.0264 was also not statistically significant (DeLong p=0.107). Model D discriminated better than Model A (Δ AUC=0.0717, 95% CI 0.0220–0.1214; DeLong p=0.005), indicating that the combined expanded model improved discrimination, whereas an incremental AUC contribution attributable specifically to pancreatic amylase was not statistically demonstrated (Table 4).

Table 5. Comparison of complete cases included in the expanded modeling and patients excluded because of missing model data

Variable	Complete cases (n=396)	Excluded (n=737)	p
Age, mean (SD), years	38.5 (10.0)	39.4 (11.1)	0.176
Female sex, n (%)	219 (55.3)	405 (55.0)	0.910
Normal radiological findings, n (%)	319 (80.6)	634 (86.0)	0.016
Normal FE-1, n (%)	328 (82.8)	633 (85.9)	0.171

Equivalent pathological-radiology proportions were 19.4% and 14.0%, and low-FE-1 proportions were 17.2% and 14.1%, respectively. FE-1: Fecal elastase-1; SD: Standard deviation.

Of the 1133 patients, 396 had complete data for all variables required for expanded modeling and 737 did not. Complete cases and excluded patients did not differ significantly in age (38.5 ± 10.0 vs. 39.4 ± 11.1 years; $p=0.176$), female sex (55.3% vs. 55.0%; $p=0.910$), or low FE-1 prevalence (17.2% vs. 14.1%; $p=0.171$). Pathological radiological findings were more frequent among complete cases than excluded patients (19.4% vs. 14.0%; $p=0.016$), indicating a potential difference in the representativeness of the complete-case subset. These comparisons are summarized in Table 5.

Discussion

Across the 1133 patients studied here, a low FE-1 value went hand in hand with abnormal imaging, an iron-deficiency profile, mildly raised liver enzymes, elevated glucose, and reduced platelet and TSH levels. Viewed together, these associations point to something broader than an isolated pancreatic abnormality — low FE-1 in this cohort appears to be associated with a wider metabolic and nutritional disturbance.

Of all variables tested, radiological abnormality carried the strongest independent association with low FE-1 ($OR=3.22$) — a result consistent with prior evidence tying fecal elastase to structural pancreatic pathology. In a series of 40 patients with low FE-1 who went on to endoscopic ultrasound, 82.5% received a new pancreatic diagnosis, chronic pancreatitis being the most frequent [6], reinforcing the connection between abnormal elastase and demonstrable structural disease. That said, Model 1's modest AUC (0.635) indicates age, sex, and imaging together predict low FE-1 only weakly, in keeping with the observation that milder exocrine dysfunction does not always produce visible structural change.

Lower iron and higher UIBC in the low FE-1 group showed a pattern suggestive of iron deficiency and may represent an early nutritional warning sign; the 2021 UK consensus guidelines on pancreatic exocrine insufficiency specifically recommend micronutrient screening, including iron status, as part of routine assessment [7]. A recent study in heart-failure patients similarly found EPI to be independently associated with broader micronutrient depletion, reinforcing that impaired exocrine function can signal a systemic nutritional problem rather than a purely pancreatic one [8]. Elevated glucose in the abnormal group mirrors the well-documented interplay between exocrine and endocrine pancreatic function: multicenter screening data reveal a substantially higher rate of EPI among people with diabetes [5], and a UK single-center series

similarly counted diabetes mellitus among the extra-pancreatic conditions linked to EPI [9]. Among diabetic patients specifically, fecal elastase has been shown to track positively with serum lipase and C-peptide, tying exocrine status to residual endocrine function at the biomarker level [10]. Further support comes from a cross-sectional study conducted in people with metabolic-associated fatty liver disease, which reported higher glycated hemoglobin among those with EPI along with a further drop in FE-1 among patients who also had diabetes [11]. Because our study is retrospective and cross-sectional, the direction of causality between glucose elevation and reduced elastase cannot be established.

The modest yet significant rise in ALT, AST, and GGT among patients with low FE-1 deserves attention and could reflect how tightly liver and pancreatic physiology are intertwined, both structurally and functionally. A 2020 investigation correlating fecal elastase with liver biopsy findings and pancreatic fat fraction in non-alcoholic fatty liver disease likewise identified exocrine insufficiency as common regardless of diabetes status, proposing a two-way metabolic connection linking the organs [12]. A more recent cross-sectional analysis in metabolic-associated fatty liver disease echoed this, documenting lower FE-1 together with higher ALT, GGT, and alkaline phosphatase, EPI in 84.3% of the sample, and strong discrimination ($AUC=0.873$) [9]. Mechanistic exploration fell outside our study's scope, but the co-occurrence of raised liver enzymes and depressed FE-1 that we observed fits comfortably within this body of evidence. Pancreatic amylase was independently associated with low FE-1 in the expanded multivariable model. However, in same-sample comparisons, adding pancreatic amylase produced only modest increases in discrimination, which did not reach statistical significance. Thus, its independent association should not be interpreted as demonstrating a significant incremental contribution to model discrimination. This finding parallels a recent multicenter investigation of chronic and autoimmune pancreatitis patients, which concluded that while normal p-amylase and lipase levels do not exclude exocrine insufficiency, reduced values in either enzyme warrant further work-up for coexisting EPI [13].

Practically speaking, these results argue for a multi-parameter strategy when deciding which patients merit further pancreatic work-up. Current expert guidance stresses that no single non-invasive marker reliably captures mild-to-moderate EPI on its own, and instead recommends raising clinical suspicion whenever several signals converge — unexplained iron deficiency, unexplained transaminase elevation, new or worsening hyperglyce-

mia, and imaging abnormalities — rather than relying on FE-1 alone [14]. Although the expanded model showed greater discrimination than the basic model, these original models were fitted in different patient samples and therefore their AUC difference should not be interpreted as an incremental effect. These routinely available variables may help identify patients in whom further pancreatic evaluation could be considered; however, their clinical utility requires prospective validation.

A number of limitations warrant mention. Because the design was retrospective and confined to one center, causal conclusions cannot be drawn. Model 2 relied on a complete-case subset ($n=396$) representing only around a third of the full cohort, which could restrict generalizability and introduce selection bias. Imaging modality and the specific criteria used for radiological assessment were not specified in the dataset, so that particular result merits cautious interpretation. Lastly, FE-1 was measured only once per patient and was not verified against a reference pancreatic function test; confirming these findings will require prospective work across multiple centers.

Conclusion

Taken together, this large-cohort analysis links abnormal FE-1 to pathological imaging, a pattern suggestive of iron deficiency, mild liver enzyme elevation, higher glucose, and — once multivariable adjustment was applied — higher pancreatic amylase. Pancreatic amylase was independently associated with low FE-1 in the expanded multivariable model. However, its addition resulted in only modest, statistically non-significant improvements in discrimination in same-sample comparisons. These findings support an association between pancreatic amylase and low FE-1 but do not establish a significant incremental predictive contribution. Prospective validation is needed to determine whether these findings may aid the clinical evaluation of patients with suspected pancreatic exocrine dysfunction.

Disclosures

Ethics Committee Approval: The study was approved by the Istanbul Medipol University Non-Interventional Clinical Research Ethics Committee (no: 485, date: 26/03/2026).

Informed Consent: The requirement for informed consent was waived by the ethics committee because this retrospective study used existing medical records.

Conflict of Interest Statement: None declared.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

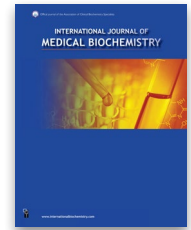
Use of AI for Writing Assistance: None declared.

Authorship Contributions: Concept – N.H.C., G.U., T.Y.; Design – N.H.C., G.U., T.Y.; Supervision – N.H.C., G.U., P.Y., T.Y., G.C.; Funding – G.U., G.C.; Materials – G.U., G.C.; Data collection and/or processing – N.H.C., G.U., P.Y.; Data analysis and/or interpretation – N.H.C., G.U., P.Y.; Literature search – N.H.C.; Writing – N.H.C., G.U., P.Y., T.Y., G.C.; Critical review – N.H.C., G.U., P.Y., T.Y., G.C.

Peer-review: Externally peer-reviewed.

References

- Capurso G, Traini M, Piciocchi M, Signoretti M, Arcidiacono PG. Exocrine pancreatic insufficiency: prevalence, diagnosis, and management. *Clin Exp Gastroenterol* 2019;12:129–139. [\[CrossRef\]](#)
- Lindkvist B. Diagnosis and treatment of pancreatic exocrine insufficiency. *World J Gastroenterol* 2013;19(42):7258–66. [\[CrossRef\]](#)
- de la Iglesia D, Agudo-Castillo B, Galego-Fernández M, Rama-Fernández A, Domínguez-Muñoz JE. Diagnostic accuracy of fecal elastase-1 test for pancreatic exocrine insufficiency: a systematic review and meta-analysis. *United European Gastroenterol J* 2025;13(8):1571–82. [\[CrossRef\]](#)
- Vanga RR, Tansel A, Sidiq S, El-Serag HB, Othman MO. Diagnostic performance of measurement of fecal elastase-1 in detection of exocrine pancreatic insufficiency: systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 2018;16(8):1220–8.e4. [\[CrossRef\]](#)
- Nunes AC, Pontes JM, Rosa A, Gomes L, Carvalheiro M, Freitas D. Screening for pancreatic exocrine insufficiency in patients with diabetes mellitus. *Am J Gastroenterol* 2003;98(12):2672–5. [\[CrossRef\]](#)
- Shobassy M, Husainat N, Tabash A, Patel K, El-Serag HB, Othman MO. Endoscopic ultrasound findings in patients diagnosed with exocrine pancreatic insufficiency by low fecal elastase-1. *Gastroenterol Res Pract* 2019;2019:5290642. [\[CrossRef\]](#)
- Phillips ME, Hopper AD, Leeds JS, Roberts KJ, McGeeney L, Duggan SN, et al. Consensus for the management of pancreatic exocrine insufficiency: UK practical guidelines. *BMJ Open Gastroenterol* 2021;8(1). [\[CrossRef\]](#)
- Vijver MAT, Bomer N, Verdonk RC, van der Meer P, van Veldhuisen DJ, Dams OC. Micronutrient deficiencies in heart failure and relationship with exocrine pancreatic insufficiency. *Nutrients* 2024;17(1):56. [\[CrossRef\]](#)
- Shandro BM, Chen J, Ritehnia J, Poullis A. Associations with pancreatic exocrine insufficiency: An United Kingdom single-centre study. *World J Clin Cases* 2021;9(31):9469–80. [\[CrossRef\]](#)
- Kangrga RN, Ignjatović SD, Dragašević MM, Jovičić SŽ, Majkić-Singh NT. Pancreatic elastase levels in feces as a marker of exocrine pancreatic function in patients with diabetes mellitus. *Lab Med* 2016;47(2):140–8. [\[CrossRef\]](#)
- Hammam GA, Abdelmalek MO, Mahmoud AA, Abdelghani M. Evidence of pancreatic exocrine insufficiency in patients with metabolic associated fatty liver disease patients. *Egypt Liver J* 2026;16:8. [\[CrossRef\]](#)
- Boga S, Koksar AR, Sen İ, Kurul Yeniay M, Yilmaz Ozguven MB, Serin E, et al. Liver and pancreas: 'Castor and Pollux' regarding the relationship between hepatic steatosis and pancreas exocrine insufficiency. *Pancreatol* 2020;20(5):880–6. [\[CrossRef\]](#)
- von Beckerath V, Klaiber U, Hackert T, Maisonneuve P, Löhr JM. Low serum pancreatic enzymes as measurement for pancreatic exocrine insufficiency in patients with chronic and autoimmune pancreatitis. *Clinical Nutrition Open Science* 2026;67:100668. [\[CrossRef\]](#)
- Whitcomb DC, Buchner AM, Forsmark CE. AGA clinical practice update on the epidemiology, evaluation, and management of exocrine pancreatic insufficiency: expert review. *Gastroenterology* 2023;165(5):1292–301. [\[CrossRef\]](#)



Research Article

Comparability of adrenocorticotrophic hormone measured in spray-dried K2-EDTA and spray-dried K3-EDTA tubes

Yasemin Ustundag, Muhammed Said Karaca, Kagan Huysal, Furkan Erdogan

Department of Clinical Biochemistry, University of Health Sciences, Bursa Yuksek Ihtisas Training and Research Hospital, Bursa, Türkiye

Abstract

Objectives: Blood samples for adrenocorticotrophic hormone (ACTH) tests are collected in dipotassium or tripotassium ethylenediaminetetraacetic acid (K2-EDTA or K3-EDTA) tubes. The interchangeable use of spray-dried K2-EDTA and K3-EDTA tubes may occasionally occur in practice and is appropriate only if both tube types yield comparable analytical results. Therefore, this study aimed to evaluate the interchangeability of these tubes for ACTH measurement.

Methods: Seventy-two volunteers were included in the study. From each participant, K2-EDTA and K3-EDTA blood samples were collected simultaneously. ACTH was measured by chemiluminescent immunoassay on a Maglumi X3 analyzer (SNIBE, Shenzhen, China). The Wilcoxon signed-rank test, Passing-Bablok regression, Mountain plot, and Bland-Altman analyses were used to evaluate differences between measurements.

Results: The median ACTH concentration in K2-EDTA tubes was 35.1 pg/mL (10.35–107.16) [7.73 pmol/L (2.28–23.60)] and in K3-EDTA tubes was 32.96 pg/mL (9.85–99.90) [7.26 pmol/L (2.17–22.00)]. The Wilcoxon signed-rank test ($z=-7.14$, $p<0.001$) demonstrated a statistically significant difference between the results. In the Bland-Altman analysis, the mean difference between the two samples was determined to be 1.49 pg/mL (0.33 pmol/L). Lin's concordance correlation coefficient was 0.989 (95% CI: 0.984–0.993), indicating good agreement between ACTH results obtained from the two tubes. Mountain plot analysis demonstrated that most differences clustered around zero, with an asymmetric right tail suggesting the presence of larger differences in a limited number of samples.

Conclusion: ACTH concentrations measured in both tubes are proportional, and the results obtained from K2-EDTA and K3-EDTA tubes support analytical comparability within the ACTH concentration range evaluated in this study. However, complete interchangeability between the results of the two tube types cannot be assumed in all clinical scenarios; further studies are needed in subjects with higher ACTH levels.

Keywords: Adrenocorticotrophic hormone, bias, K2-EDTA; K3-EDTA, pre-analytical phase, test tube comparison

How to cite this article: Ustundag Y, Karaca MS, Huysal K, Erdogan F. Comparability of adrenocorticotrophic hormone measured in spray-dried K2-EDTA and spray-dried K3-EDTA tubes. Int J Med Biochem 2026;9(4):217–222.

Human adrenocorticotrophic hormone (ACTH) is the stimulator of glucocorticoid synthesis and secretion by the adrenal cortex. Accurate measurement of the plasma ACTH concentration is crucial for diagnosing disorders of the hypothalamic–pituitary–adrenal axis, including Cushing's syndrome and adrenal insufficiency [1, 2]. Meanwhile, pre-analytical factors such as sample type and storage time have long been recognized as major determinants of ACTH assay reliability [3–7].

Among the pre-analytical variables, the blood collection process, including tube selection, plays a critical role [6, 8]. The type and concentration of additives in blood collection tubes can significantly influence test results by affecting sample stability, clotting, or analyte concentration. Therefore, proper selection and validation of blood collection tubes are essential steps in minimizing pre-analytical variability and ensuring accurate and reliable laboratory outcomes [8, 9].

Address for correspondence: Yasemin Ustundag, MD. Department of Clinical Biochemistry, University of Health Sciences, Bursa Yuksek Ihtisas Training and Research Hospital, Bursa, Türkiye

Phone: +90 532 482 36 92 **E-mail:** yaseminustundagyeah@gmail.com **ORCID:** 0000-0003-2415-0372

Submitted: May 22, 2026 **Accepted:** August 18, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



In our laboratory, ACTH concentrations are measured using a chemiluminescent immunoassay method on a Maglumi X3 analyzer (SNIBE, Shenzhen, China). Ethylenediaminetetraacetic acid (EDTA) tubes are reported by the manufacturer as the preferred anticoagulant according to the ACTH analysis sheet. EDTA is present in blood collection tubes as the dipotassium (K2), tripotassium (K3), or sodium EDTA salts; however, no major international guideline specifically recommends K2-EDTA over K3-EDTA, and both tube types are considered acceptable for ACTH measurement [10].

Our laboratory is a large, multi-site facility formed by the merger of laboratories operating at different locations. The interchangeable use of K2-EDTA and K3-EDTA tubes may occasionally occur in routine practice and is appropriate only if both tube types yield comparable analytical results.

Therefore, this study aimed to evaluate the interchangeability of these tubes for ACTH measurement.

Materials and Methods

This study was conducted at a University Hospital. The study protocol was approved by the Bursa Yüksek İhtisas Training and Research Hospital Ethics Committee (Date: 11.03.2026, No: 2024-TBEK 2026/03-03) and carried out in accordance with the Declaration of Helsinki [11]. Written informed consent was obtained from all participants.

The study included healthcare workers and patients presenting to the endocrinology outpatient clinic. In accordance with the recommendations of the Clinical and Laboratory Standards Institute (CLSI) EP09 guideline, 72 participants (36 females, 36 males) aged 18–65 years were included in the study, and no formal sample size calculation was performed (Table 1) [12].

Participants with acute illness, pregnancy, or recent steroid use were excluded.

Venous blood samples were collected from participants between 9:00 a.m. and 10:00 a.m. in the biochemistry laboratory to avoid potential pre-analytical variations related to sample transportation.

Manufacturer instructions for the ACTH assay generally allow the use of both K2-EDTA and K3-EDTA tubes. From each participant, blood was collected consecutively into K2-EDTA (Lot: 3342.0012.24) and K3-EDTA (Lot: 3104.0010.24) tubes (Vacu-sera; Disera, İzmir, Türkiye), with no predetermined order of tube collection.

The inner wall of the tubes is coated with spray-dried K2 or K3 EDTA rather than a liquid form, which minimizes dilution and maintains the original blood composition, ensuring test accuracy at room temperature (between 22–24°C in our laboratory during the study period). The tubes were filled with approximately 2 mL of venous blood (as recommended by the manufacturer), thus ensuring adequate tube filling and an appropriate blood-to-anticoagulant ratio (1.8 mg per 1 mL of blood).

After specimen collection, we mixed tube contents by inverting them 10 times and centrifuged them at 1500 × g for 10 min (2–8°C) immediately after blood collection. ACTH levels were analyzed within 30 minutes. No hemolysis or lipemia was observed in any of the samples, as assessed objectively using the automated hemolysis and lipaemia indices.

For paired ACTH test results, a maximum allowable bias of ±9.08 pg/mL (±2.0 pmol/L) for concentrations ≤90.82 pg/mL (≤20 pmol/L), or ±10% for concentrations >90.82 pg/mL (>20 pmol/L), was selected based on the Royal College of Pathologists of Australasia (RCPA) allowable limits of performance

Table 1. General characteristics of the study participants and summary of method comparison statistics for ACTH measurements in K2-EDTA and K3-EDTA tubes

Variable	Result
Number of participants	72
Female/Male, n	36/36
Age mean±SD, (range)	29.8±9.4 (18–65)
ACTH concentration, pg/mL (pmol/L), median (range)	
K ₂ -EDTA tube	35.1 (10.35-107.16) [7.73 (2.28–23.60)]
K ₃ -EDTA tube	32.96 (9.85-99.90) [7.26 (2.17–22.00)]
Wilcoxon signed-rank test	z=-7.14; p<0.001
Paired samples with higher ACTH in K ₂ -EDTA, n (%)	69 (95.8)
Passing–Bablok regression analysis	
Regression equation	y=0.086(-0.06 to 0.18) + 1.01×(0.99 to 1.03)
Pearson correlation coefficient	r=0.97; p<0.001
Cusum test for linearity	p=0.68
Agreement analysis	
Mean bias (Bland–Altman)	1.49 pg/mL (0.33 pmol/L)
Spearman correlation analysis	ρ=0.166; p=0.164
Lin's concordance correlation coefficient	0.989 (95% CI: 0.984–0.993)

ACTH: Adrenocorticotrophic Hormone; K₂-EDTA: Dipotassium ethylene diamine tetra-acetic acid; K₃-EDTA: Tripotassium ethylene diamine tetra-acetic acid.

for ACTH [13]. This hybrid allowable bias criterion enables a clinically appropriate assessment of analytical performance for ACTH across the measurement range to minimize laboratory-related diagnostic errors.

Plasma ACTH concentrations were analyzed using a commercially available sandwich chemiluminescence immunoassay (Maglumi ACTH kit, SNIBE, Shenzhen, China) on a Maglumi X3 (SNIBE, Shenzhen, China) analyzer. Manufacturer-provided reagents, calibrators, and internal quality control materials were used. Quality control materials at two concentrations [35 and 200 pg/mL (7.71 and 44.04 pmol/L)] were analyzed daily in duplicate for 20 days. Our laboratory participates in the Serocon External Quality Control Program (Konya, Türkiye) during the study interval.

Statistical Analysis

MedCalc statistical software (MedCalc, Mariakerke, Belgium) was used to analyze the data. Data normality was tested using the Kolmogorov–Smirnov test. Agreement and differences between ACTH concentrations measured in K2-EDTA and K3-EDTA tubes were analyzed using the Wilcoxon signed-rank test, Passing–Bablok regression, Spearman correlation analysis, Mountain plot (folded empirical cumulative distribution plot), and Bland–Altman analysis. The strength of agreement between the two results was also assessed using Lin’s concordance correlation coefficient (CCC), which incorporates both precision and accuracy [14]. Statistical significance was defined as $p < 0.05$.

Results

In day-to-day quality control assessments, the coefficient of variation (CV) for ACTH measurements was 3.99% for QC1 [35 pg/mL; range 24.5–45.5 (7.71 pmol/L; range 5.4–10.01)] and 2.67% for QC2 [200 pg/mL; range 140–260 (44.04 pmol/L; range 30.83–57.26)].

The median ACTH concentration from 72 participants in K2-EDTA tubes was 35.1 pg/mL (10.35–107.16) [7.73 pmol/L (2.28–23.60)] and in K3-EDTA tubes was 32.96 pg/mL (9.85–99.90) [7.26 pmol/L (2.17–22.00)]. The Wilcoxon signed-rank test ($z = -7.14$, $p < 0.001$) demonstrated a statistically significant difference between the two tubes, showing a systematic positive bias for ACTH results from K2-EDTA tubes. In this study, the ACTH result obtained from 69 K2-EDTA tubes was higher than the result obtained from K3-EDTA tubes.

The regression line equation [$y = 0.086 + 1.01x$; (95% CI for intercept -0.06 to 0.18 and for slope 0.99 to 1.03)] and Pearson correlation coefficient ($r = 0.97$, $p < 0.001$) indicated a strong relationship between ACTH results from the two tubes (Fig. 1). The Cusum test for linearity ($p = 0.68$) showed no significant deviation from linearity (Table 1). The 95% CI for the intercept includes zero, indicating that there is a small constant difference between the results obtained from the two specimen collection tube types. The 95% CI for the slope includes a value of one, showing no proportional difference between paired ACTH results and indicating that the measurement bias did not increase systematically with increasing ACTH concentrations [15].

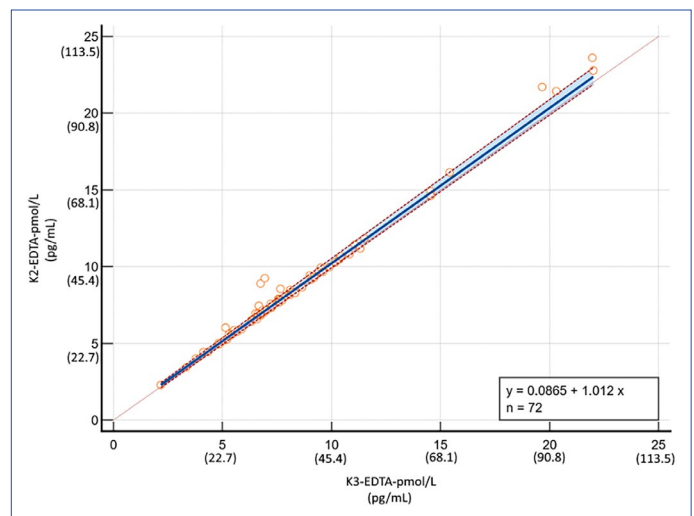


Figure 1. Passing-Bablok regression analysis of ACTH concentrations measured.

Passing-Bablok regression analysis of the comparability of ACTH concentrations measured in K2-ethylenediaminetetraacetic acid (EDTA) and K3-EDTA Vacusera tubes (Disera, İzmir, Türkiye). Solid line represents regression line; dashed lines represent confidence interval (CI) for the regression line; dotted line is the line of identity ($x = y$). $n = 72$; concentration range 9.85–107.16 pg/mL (2.17–23.60 pmol/L); Pearson correlation coefficient $r = 0.97$, $p < 0.001$. Regression line equation: $y = 0.086 + 1.01x$; 95% CI for intercept -0.06 to 0.18 and for slope 0.99 to 1.03 indicated good agreement.

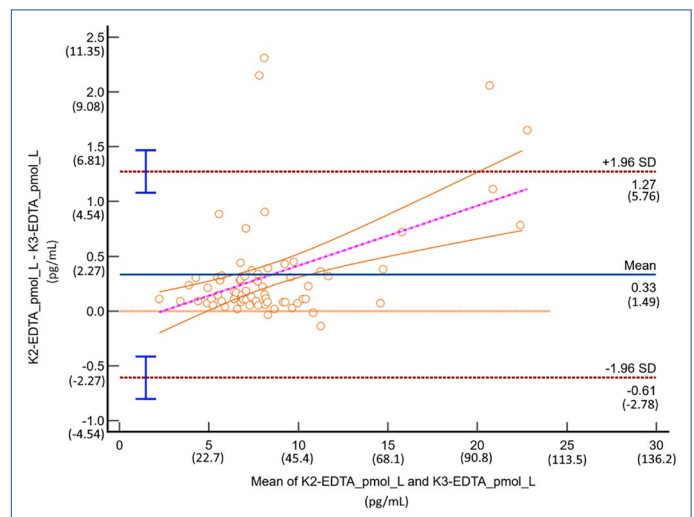


Figure 2. Bland-Altman plot of the comparability of ACTH measurements.

Solid lines represent mean difference and the outer dotted lines represent limits of agreement (mean difference ± 1.96 SD). K2-EDTA: Dipotassium ethylene diamine tetra-acetic acid tubes K3-EDTA, tripotassium ethylene diamine tetra-acetic acid tubes. ACTH: Human adrenocorticotrophic hormone.

Bland–Altman analysis demonstrated a mean difference of 1.49 pg/mL (0.33 pmol/L) between the ACTH results obtained from K2-EDTA and K3-EDTA tubes (Table 1). This bias is lower than the maximum allowable bias of ± 9.08 pg/mL (2.0 pmol/L), indicating acceptable agreement for ACTH concentrations between 9.85 and 107.16 pg/mL (2.17 and 23.60 pmol/L) (Fig. 2). However, the analysis showed that the two measurements do

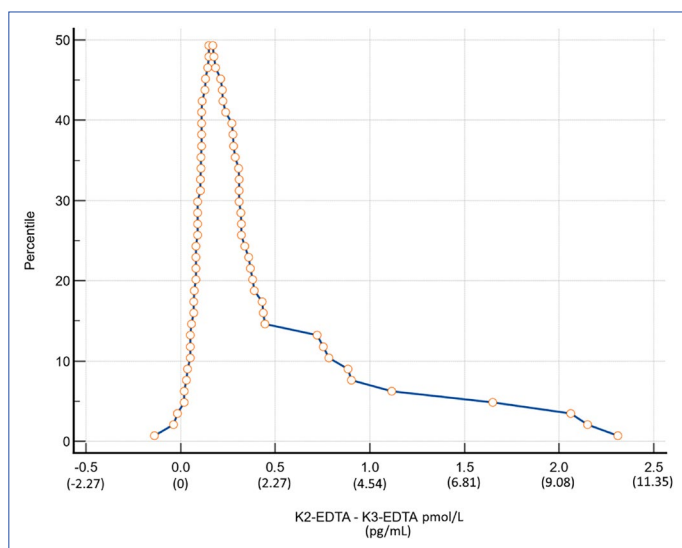


Figure 3. A mountain plot (or "folded empirical cumulative distribution plot") of the ACTH measurements.

Percentile for each ranked difference between ACTH measurements of K2-EDTA tubes and K3-EDTA tubes. K2-EDTA: Dipotassium ethylene diamine tetra-acetic acid tubes K3-EDTA, tripotassium ethylene diamine tetra-acetic acid tubes. ACTH: Human adrenocorticotrophic hormone.

not always agree, and some measurements lay outside the limits of agreement. Spearman correlation analysis showed no statistically significant association between the measurement difference and the mean ACTH concentration ($p=0.166$, $p=0.164$), indicating that the difference between the two tube types did not increase significantly with increasing ACTH concentrations across the studied range.

Lin's concordance correlation coefficient was 0.989 (95% CI: 0.984–0.993), demonstrating good agreement between the ACTH results obtained from K2-EDTA and K3-EDTA tubes; however, this finding alone does not confirm interchangeability.

Mountain plot analysis demonstrated that most differences clustered around zero, indicating minimal overall bias (Fig. 3). However, the asymmetric right tail suggests the presence of larger differences in a limited number of samples, which may be clinically relevant near decision limits.

Discussion

In the present study, plasma ACTH concentrations in blood samples collected in spray-dried K2-EDTA tubes were slightly higher than those collected in spray-dried K3-EDTA tubes. To the best of our knowledge, no study has yet investigated the comparability of specimens collected in K2-EDTA and K3-EDTA tubes for ACTH measurement. Chang et al. [16] compared the plasma levels of α -synuclein and neurofilament light chain collected in K2- and K3-EDTA tubes and found no clinically significant differences that influenced the measurement of these biomarkers.

Leatham et al. [17] demonstrated that the results of routine blood bank tests performed on 102 blood donor samples and 100 patient samples taken with K2-EDTA and K3-EDTA were

equivalent. These results are similar to our finding of close analytical agreement between the two tube types for ACTH measurement within the evaluated concentration range. However, when evaluating disorders of the hypothalamic–pituitary–adrenal axis, ACTH concentrations are often interpreted using clinical decision cut-off points. Consequently, small analytical differences close to these limits can lead to diagnostic variation.

Each molecule of K2-EDTA and K3-EDTA contains two and three dissociable potassium ions, respectively. K3-EDTA contains more potassium per volume, leading to a slightly greater dilution effect, which may explain the lower ACTH levels measured in these tubes [18].

When clinical decisions are based on fixed cut-off values, variability between results becomes especially important.

A maximum allowable bias of ± 9.08 pg/mL (for values ≤ 90.82 pg/mL) [± 2.0 pmol/L (for values ≤ 20 pmol/L)] or $\pm 10\%$ [for values > 90.82 pg/mL (> 20 pmol/L)] between K2-EDTA and K3-EDTA plasma ACTH measurements was adopted as recommended by the Royal College of Pathologists of Australasia allowable error guidelines [13]. Following these performance limits confirms the acceptable analytical agreement between the results obtained from both tube types, demonstrating that ACTH measurements remain consistent within the studied concentration range. In our study, only 4 volunteers had results greater than 90.82 pg/mL (20 pmol/L). Moreover, in all volunteers, the observed differences in results between these tubes were lower than this predefined criterion.

Bland–Altman analysis demonstrated that the mean difference between the results from the K2-EDTA and K3-EDTA tubes was 1.49 pg/mL (0.33 pmol/L), and the bias exhibited by either method did not exceed the acceptable level of inaccuracy [13].

The combined use of mountain plot visualization and Lin's concordance correlation coefficient provides a robust evaluation of the agreement between the two ACTH measurements. Although the paired results demonstrate good overall agreement, complete interchangeability across all clinical scenarios cannot be assumed [14].

Toprak et al. [19] found that increasing EDTA concentration in blood collection tubes causes a decrease in ACTH concentration. Although their study does not clearly distinguish between K2- and K3-EDTA tubes, the effect of EDTA salt amount is emphasized. They stated that under-filled EDTA tubes should be considered unsuitable for ACTH measurement.

Fraissinet et al. [20] compared ACTH stability in blood collected in K2-EDTA and K3-EDTA+aprotinin tubes. The ACTH concentration in K3-EDTA tubes was higher, which could be explained by the effect of aprotinin rather than the tube type itself. They concluded that pre-analytical factors have a greater impact on ACTH stability than the collection tube type.

Meanwhile, the difference in anticoagulant composition may influence peptide stability, highlighting the importance of standardizing tube selection across laboratories.

Limitations

This study has several limitations. The number of participants was relatively small ($n=72$), and ACTH concentrations in most volunteers were within the reference interval, with only four participants having ACTH concentrations >90.82 pg/mL (>20 pmol/L). Therefore, our findings may not be generalizable to samples with markedly elevated ACTH concentrations, adrenal insufficiency, Cushing's syndrome, ectopic ACTH syndrome, or samples with values near clinically important decision limits.

Conclusion

In conclusion, ACTH concentrations measured in both tubes are proportional, and the results obtained from K2-EDTA and K3-EDTA tubes support analytical comparability within the ACTH concentration range evaluated in this study.

The limited representation of the higher concentration range also restricts conclusions regarding assay agreement at clinically important decision points, where even small analytical differences between specimen types may have disproportionate clinical consequences, including in the evaluation of ACTH-dependent disorders. Further studies including larger cohorts with higher ACTH concentrations are warranted to confirm these findings. These studies should also address the importance of implementing standardized pre-analytical protocols to ensure reliable analytical results.

Disclosures

Ethics Committee Approval: The study was approved by the Bursa Yüksek İhtisas Training and Research Hospital Clinical Research Ethics Committee (no: 2024-TBEK 2026/03-03, date: 11/03/2026).

Informed Consent: Written informed consent was obtained from all participants.

Conflict of Interest Statement: None declared.

Funding: The author declares that this study has received no financial support.

Use of AI for Writing Assistance: None declared.

Authorship Contributions: Concept – Y.U., M.S.K., K.H., F.E.; Design – Y.U., M.S.K., K.H., F.E.; Supervision – Y.U., M.S.K., K.H., F.E.; Funding – Y.U., M.S.K., K.H., F.E.; Materials – Y.U., M.S.K., K.H., F.E.; Data collection and/or processing – Y.U., M.S.K., K.H., F.E.; Data analysis and/or interpretation – Y.U., M.S.K., K.H., F.E.; Literature search – Y.U., M.S.K., K.H., F.E.; Writing – Y.U., M.S.K., K.H., F.E.; Critical review – Y.U., M.S.K., K.H., F.E.

Peer-review: Externally peer-reviewed.

References

1. Fleseriu M, Auchus R, Bancos I, Ben-Shlomo A, Bertherat J, Birmasz NR, et al. Consensus on diagnosis and management of Cushing's disease: a guideline update. *Lancet Diabetes Endocrinol* 2021;9:847–75. [CrossRef]
2. Flowers KC, Shipman KE. Pitfalls in the diagnosis and management of hypercortisolism (Cushing syndrome) in humans; A review of the laboratory medicine perspective. *Diagnostics (Basel)* 2023;13(8):1415. [CrossRef]
3. Turkon H, Toprak B, Yalcin H, Colak A, Ozturk N. The Effectiveness of temperature versus aprotinin in maintaining the preanalytical stability of adrenocorticotrophin. *Lab Med* 2016;47(4):279–82. [CrossRef]
4. Lin Y, Spies NC, Zohner K, McCoy D, Zaydman MA, Farnsworth CW. Pre-analytical phase errors constitute the vast majority of errors in clinical laboratory testing. *Clin Chem Lab Med* 2025;63:1709–15. [CrossRef]
5. Dib A, Leleu D, Lemaire S, Duvillard L, Ménégaut L, Denimal D. Is the stability of ACTH in whole blood a genuine concern during the preanalytical phase? A systematic review. *Biochem Med (Zagreb)* 2025;35(1):010502. [CrossRef]
6. Seo JD, Lee YJ, Ha C, Choi JH, Shin SH, Han HY, et al. Comparison of three blood collection tubes for 16 biochemical analytes and stability assessment of selected analytes: VACUETTE® CAT serum Sep clot activator tube, VACUETTE® LH lithium heparin Sep tube, and VACUETTE® CAT serum fast separator tube. *Ann Clin Biochem* 2025;62(5):395–407. [CrossRef]
7. Hillebrand JJ, Zhou L, Marcinkus MA, Datwyler M, Gawel SH, Martens F, et al. Instability of corticotropin during long-term storage - myth or reality? *Clin Chem Lab Med* 2021;60(1):60–5. [CrossRef]
8. Budak YU, Huysal K, Bulut M, Polat M. Evaluation in an emergency department of rapid separator tubes containing thrombin for serum preparation prior to hs-cTnT and CK-MB analyses. *BMC Clin Pathol* 2013;13:20. [CrossRef]
9. Ahn S, Cho SM, Shin H, Lee KA. Comparison of improvacuter EDTA tube with BD vacutainer EDTA tube for routine hematological analysis: Clinical significance of differences, stability study, and effects of K₂ and K₃ EDTA. *J Lab Med Qual Assur* 2016;38(2):77–86. [CrossRef]
10. Banfi G, Salvagno GL, Lippi G. The role of ethylenediamine tetraacetic acid (EDTA) as *in vitro* anticoagulant for diagnostic purposes. *Clin Chem Lab Med* 2007;45(5):565–76. [CrossRef]
11. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310(20):2191–4. [CrossRef]
12. Budd JR, Durham AP, Gwise TE, Hawkins DM, Holland M, Iriarte G, et al. Measurement procedure comparison and bias estimation using patient samples. Wayne (PA): Clinical and Laboratory Standards Institute; 2018. CLSI guideline EP09c.
13. Westgard S. RCPA Allowable limits of performance for biochemistry - Westgard QC. 2016; Available at: <https://westgard.com/clia-a-quality/quality-requirements/rcpa-biochemistry.html> Accessed: September 16, 2026.
14. Lin L, Hedayat AS, Sinha B, Yang M. Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association* 2002;97(457):257–70. [CrossRef]
15. Bilić-Zulle L. Comparison of methods: passing and bablok regression. *Biochem Med (Zagreb)* 2011;21(1):49–52. [CrossRef]
16. Chang KH, Liu KC, Lai CS, Yang SY, Chen CM. Assessing plasma levels of α -synuclein and neurofilament light chain by different blood preparation methods. *Front Aging Neurosci* 2021;13:759182. [CrossRef]

17. Leathem S, Zantek ND, Kemper M, Korte L, Langeberg A, Sandler SG. Equivalence of spray-dried K2EDTA, spray-dried K3EDTA, and liquid K3EDTA anticoagulated blood samples for routine blood center or transfusion service testing. *Immunohematology* 2003;19(4):117–21. [\[CrossRef\]](#)
18. Lorde N, Mahapatra S, Kalaria T, Gama R. Potassium ethylenediaminetetraacetic acid (kEDTA) sample cross-contamination: prevalence, consequences, identification, mechanisms, prevention and mitigation. *J Lab Precis Med* 2024;9:34. [\[CrossRef\]](#)
19. Toprak B, Yalcin H, Ari E, Colak A. EDTA interference in electrochemiluminescence ACTH assay. *Ann Clin Biochem* 2016;53(6):699–701. [\[CrossRef\]](#)
20. Fraissinet F, Girot H, Gillibert A, Melin A, Fettig J, Brunel V. Stability of adrenocorticotrophic hormone in whole blood samples: effects of storage conditions. *Biochem Med (Zagreb)* 2024;34(3):030702. [\[CrossRef\]](#)



Research Article

End-user verification of the Maccura F810 hematology analyzer: A comparison with Mindray BC-6200

 Havva Yasemin Cinpolat,  Muserref Hilal Sehitoglu Bingol

Department of Medical Biochemistry, Canakkale Onsekiz Mart University Faculty of Medicine, Canakkale, Türkiye

Abstract

Objectives: Analytical performance verification is essential before implementing hematology analyzers in routine clinical practice. This study aimed to evaluate the performance characteristics of the Maccura F810 hematology analyzer.

Methods: This study was conducted at the Medical Biochemistry Laboratory of Çanakkale Onsekiz Mart University Hospital between January and April 2023. Analytical performance was verified by evaluating precision (within-run and between-batch), carry-over, and linearity in accordance with the Clinical and Laboratory Standards Institute and the International Council for Standardization in Hematology guidelines. Additionally, 356 K2-EDTA samples from patients were used for comparative analysis with the Mindray BC-6200, including Passing-Bablok regression, Bland-Altman analysis, and bias estimation at medical decision points. We compared the results with the manufacturer's claimed performance specifications and further evaluated them using state-of-the-art (SOTA) and minimum biological variation (MBV) performance criteria.

Results: All precision results met the manufacturer's performance specifications, whereas only monocyte, eosinophil, and basophil counts exceeded the SOTA precision limits at certain concentration levels. Carry-over was negligible for white blood cell count, red blood cell count, hemoglobin, and platelet count. Linearity was confirmed across the concentration ranges specified by the manufacturer. Bland-Altman analysis showed agreement with the manufacturer's specifications and SOTA criteria for all parameters, while all parameters except monocyte count also fulfilled the MBV criteria. Predicted bias at predefined medical decision points was within the MBV and SOTA limits for all evaluated parameters, except for platelet count at $50 \times 10^9/L$, where the predicted bias was 8.00% (95% CI, 3.21%-12.89%), exceeding both the MBV (7.5%) and SOTA (6.4%) limits.

Conclusion: Overall, the Maccura F810 demonstrated satisfactory analytical performance and is suitable for routine clinical laboratory use, despite higher CVs observed for a limited number of parameters and a predicted PLT bias at $50 \times 10^9/L$ that exceeded both acceptance criteria.

Keywords: Complete blood count, hematology, instrumental analysis, method verification, performance evaluation

How to cite this article: Cinpolat HY, Sehitoglu Bingol MH. End-user verification of the Maccura F810 hematology analyzer: A comparison with Mindray BC-6200. Int J Med Biochem 2026;9(4):223–231.

The complete blood count (CBC) is one of the most commonly requested hematological tests and provides essential information for patient diagnosis and clinical management [1]. It plays a key role in the diagnosis and monitoring of hematological disorders, including anemia, hematologic malignancies, and coagulation disorders. It is also widely used to identify infections, allergic conditions, and immune system disorders

[2]. Automated hematology analyzers measure hemoglobin (HGB) concentration, quantify and characterize blood cell populations, differentiate leukocyte subpopulations, and assess cell volume and cellular content. They also calculate diagnostic indices based on these measurements. These analyses are performed simultaneously through fully automated processes, supporting rapid, standardized results [3].

Address for correspondence: Havva Yasemin Cinpolat, MD. Department of Medical Biochemistry, Canakkale Onsekiz Mart University Faculty of Medicine, Canakkale, Türkiye

Phone: +90 555 340 22 59 **E-mail:** yaseminarattan@gmail.com **ORCID:** 0000-0002-7161-2907

Submitted: July 27, 2026 **Revised:** August 25, 2026 **Accepted:** August 28, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



Technological advances have substantially enhanced the analytical performance of hematology analyzers by integrating multiple measurement technologies [4]. Historically, CBC analysis was performed manually. Automation was introduced following the description of the Coulter principle in the late 1940s, leading to the development of the first automated hematology analyzer in the 1950s [5]. Early analyzers were based on the electrical impedance principle combined with basic optical methods for cell counting and volume measurement. Subsequent technological developments enabled the assessment of intracellular structure and cellular composition. The incorporation of multi-angle light-scatter analysis made comprehensive leukocyte differentiation possible. More recently, fluorescent staining methods targeting intracellular components, particularly nucleic acids, have enabled highly accurate measurement of reticulocytes and nucleated red blood cells (NRBCs). The same technology also supports a six-part leukocyte differential by identifying immature granulocytes (IG) [6]. In addition to quantitative cell counts, analyzer-generated flags provide valuable information about peripheral blood smear abnormalities and help reduce the need for manual microscopic review [7].

Hematology analyzers play a key role in clinical decision-making, so every new device should undergo an extensive manufacturer validation process before routine use to ensure accurate and reliable performance. Following manufacturer validation, clinical laboratories are required to perform instrument verification to confirm analytical performance before patient testing. Validation and verification procedures should be conducted in accordance with the recommendations of the Clinical and Laboratory Standards Institute (CLSI), the International Council for Standardization in Haematology (ICSH), and the ISO 15189 standard [8–10]. Although these guidelines differ in certain aspects, verification typically includes the evaluation of precision, accuracy, method comparability, carry-over, linearity, limit of blank, limit of detection, reference intervals, and the analytical measuring interval [1]. However, current guidelines do not define universally accepted acceptance criteria or minimum performance requirements for individual analytical parameters [11]. Therefore, verification of the manufacturer's performance claims is necessary [1]. Establishing appropriate acceptance criteria and interpreting verification results are the responsibility of laboratory specialists, who should apply professional judgment in accordance with the analyzer's intended clinical use [12].

To the best of our knowledge, no previous verification study of the Maccura F810 hematology analyzer has been reported in the literature. Therefore, this study was designed to verify its analytical performance against the manufacturer's performance claims, state-of-the-art (SOTA) criteria, and minimum biological variation (MBV) specifications. In addition, we compared its performance with that of the Mindray BC-6200 under routine laboratory conditions [3, 13].

Materials and Methods

Analyzer

The F810 (Maccura Biotechnology, Chengdu, China) is a third-generation hematology analyzer designed for high-volume laboratories and capable of processing up to 100 samples per hour. It uses fluorescent dyes that target DNA and RNA to differentiate leukocyte subpopulations by nucleic acid content. The analyzer incorporates multiple dedicated analytical channels. The FCW channel detects white blood cells (WBCs), basophils, and NRBCs, while the FCD channel generates the leukocyte scattergram. Together, these channels provide a six-part WBC differential, including IGs, and generate 29 research parameters. Red blood cells (RBCs) and platelets (PLTs) are measured in a separate channel using the sheath-flow impedance method, whereas HGB concentration is determined by colorimetry. The FCR channel measures reticulocytes, reports six reticulocyte-related parameters, and performs optical platelet counting (PLT-O) for samples with low PLT counts. The FCP channel uses fluorescence-based platelet counting (PLT-F), while the FCA channel is designed to detect immature cells of the bone marrow and lymphoid lineage. In addition, the analyzer includes an LW mode for samples with low WBC counts and a dedicated body-fluid analysis mode.

Comparative analyzer

The operating principle and technological characteristics of the BC-6200 hematology analyzer (Mindray Bio-Medical Electronics, Shenzhen, China) have been described previously [14]. The analyzer provides a six-part WBC differential using SF Cube technology, which combines light scatter (S), fluorescence (F), and three-dimensional (3D) analysis. For leukocyte differentiation, cells are stained with a fluorescent dye according to their nucleic acid content. The WNB channel differentiates WBCs, basophils, and NRBCs. In the RBC/PLT channel, RBCs and PLTs are counted using the electrical impedance method, while HGB concentration is measured by a cyanide-free colorimetric method.

Samples

This study was approved by the Çanakkale Onsekiz Mart University Clinical Research Ethics Committee (Date: 04.01.2023, Decision no: 2023/01-18). Normal and abnormal whole-blood samples were collected from outpatient, inpatient, and intensive care unit patients during routine laboratory practice at the Medical Biochemistry Laboratory of Çanakkale Onsekiz Mart University Hospital between January and April 2023. Blood samples were collected into K2EDTA tubes (Vacusera, Disera Medical, İzmir, Türkiye) and analyzed within 2 hours of collection.

Precision

We evaluated within-run precision (repeatability) and between-batch precision in accordance with the recommendations of ICSH [8]. For repeatability, we selected patient sam-

ples representing low abnormal, clinically relevant (medical decision level), and high abnormal concentrations, and each sample was analyzed 10 consecutive times in a single analytical run. Between-batch precision was assessed using three levels of CBC-X hematology quality control materials (R&D Systems, Minneapolis, MN, USA), which were analyzed over 20 consecutive days. During the between-batch precision study, a single lot was used for each level: Lot No. 23700813 for the low level, Lot No. 23700814 for the normal level, and Lot No. 23700815 for the high level.

Carry-over

The carry-over study was performed in accordance with the ICSH guidelines [8]. For WBC, RBC, HGB, and PLT, a high-level sample was analyzed in triplicate (H1, H2, and H3), followed by a low-level sample diluted with particle-free plasma, which was also analyzed in triplicate (L1, L2, and L3). The carry-over percentage was calculated using the following formula: $[(L1 - L3) / (H3 - L3)] \times 100$. We compared the calculated values with the SOTA limits and the manufacturer's specifications [3].

Linearity

Linearity was assessed according to the procedure recommended by the ICSH guidelines. For WBC, RBC, HGB, and PLT, we prepared an 11-point serial dilution series (0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100%) from high-concentration samples using the diluent reagent. Each dilution was analyzed in triplicate, and the mean value was calculated. We performed linear regression analysis by comparing the observed and expected values [8]. The results were compared with the manufacturer's claim and MBV total allowable error (TEa) for verification. In addition, linearity was evaluated separately for low-level WBC, mid-level WBC, and low-level PLT samples. In LW mode, we also assessed linearity specifically for low-level WBC samples.

Comparison study

A comparative evaluation with the Mindray BC-6200 was performed in accordance with the CLSI EP09C guideline [15]. A total of 356 normal and abnormal whole-blood samples were analyzed on both analyzers. Method agreement was evaluated using Passing-Bablok regression analysis, Bland-Altman plots, and bias estimation at predefined medical decision points.

Statistical analysis

Statistical analyses were performed using MedCalc version 23.6.1 (MedCalc, Ostend, Belgium). Data distribution was assessed with the Kolmogorov-Smirnov test. For the precision study, the mean, standard deviation (SD), and coefficient of variation (CV) were calculated. The obtained CVs were compared with the MBV, SOTA, and manufacturer-specified performance criteria [3, 13]. Method comparison data were evaluated using Passing-Bablok regression analysis. Analytical performance was considered acceptable when the 95%

confidence intervals (CIs) for the intercept and slope included 0.0 and 1.0, respectively. If these values were not included in the corresponding CIs, the deviation was interpreted as a constant bias in the intercept and a proportional bias in the slope. Bland-Altman analysis was performed by plotting the percentage difference between the two analyzers against their mean values. Depending on the data distribution, either the mean or the median percentage difference (MD%) was calculated. Bias at clinical decision points was estimated using Passing-Bablok regression analysis, with 95% CIs generated via nonparametric bootstrap sampling. The predefined medical decision points were determined to reflect clinically important conditions, critical values, or reference intervals, including mild anemia (HGB 12 g/dL), severe anemia (HGB 8 g/dL), polycythemia vera (HGB 16.5 g/dL), severe leukopenia (WBC $2 \times 10^9/L$), leukopenia (WBC $4 \times 10^9/L$), leukocytosis (WBC $13 \times 10^9/L$), neutropenia (Neutrophils $1.8 \times 10^9/L$), microcytosis (mean corpuscular volume (MCV) 80 fL), macrocytosis (MCV 100 fL), severe thrombocytopenia (PLT $50 \times 10^9/L$), thrombocytopenia (PLT $150 \times 10^9/L$), and thrombocythemia (PLT $450 \times 10^9/L$) [16–20]. The calculated MD% values and predicted bias at the predefined medical decision points were compared with the SOTA and MBV bias limits [3, 13].

Results

Precision

The within-run and between-batch precision results are summarized in Table 1, along with the manufacturer's performance specifications, MBV, and SOTA limits. The calculated within-run and between-batch CVs for all parameters met the Maccura performance specifications. Except for the between-batch CV for MCV, the within-run CV at the normal level and the between-batch CV at the low level for monocyte count, and the within-run CVs at the normal and high levels together with the between-batch CVs at the low and normal levels for basophil count, all within-run and between-batch CVs satisfied the MBV criteria. Exceedance of the SOTA limits was observed only for the within-run CV at the normal level and the between-batch CVs at all concentration levels for monocyte count, the within-run CV at the high level for eosinophil count, and the within-run CVs at the normal and high levels for basophil count.

Carry-over

No significant carry-over was observed for WBC, RBC, HGB, or PLT. The calculated carry-over rates, together with the manufacturer's specifications and SOTA targets, are presented in Table 2.

Linearity

Linearity was evaluated over the following analytical ranges: low-level WBC, $0-1.00 \times 10^9/L$; mid-level WBC, $0-25.45 \times 10^9/L$; high-level WBC, $14.21-140.79 \times 10^9/L$; RBC, $0-7.85 \times 10^{12}/L$; HGB, $0-20.8$ g/dL; low-level PLT, $0.33-25 \times 10^9/L$; and high-level PLT,

Table 1. Within run precision using different levels of whole blood samples and between-batch precision using quality control materials

Parameter	Level	Within-run precision			Between-batch precision			Maccura	MBV	SOTA (within-run)	SOTA (between-batch)
		Mean	SD	CV%	Mean	SD	CV%				
WBC (10 ⁹ /L)	Low	0.65	0.02	3.08	3.95	0.09	2.28	–	8.3	6.0	6.0
	Normal	6.20	0.08	1.29	7.11	0.10	1.41	≤2.5		2.5	2.5
	High	21.91	0.17	0.78	20.77	0.26	1.25	≤2.5		1.5	1.5
RBC (10 ¹² /L)	Low	2.25	0.02	0.89	2.42	0.02	0.83	–	2.0	1.1	1.1
	Normal	5.47	0.04	0.73	4.56	0.03	0.66	≤1.5			
	High	8.31	0.05	0.60	5.76	0.03	0.52	≤1.5			
HGB (g/dL)	Low	5.2	0.04	0.77	6.1	0.06	0.98	–	1.3	0.9	1.0
	Normal	12.7	0.06	0.47	13.7	0.09	0.66	≤1			
	High	17.6	0.07	0.40	18.1	0.09	0.50	≤1			
MCV (fL)	Low	61	0.16	0.26	80	0.61	0.76 ^a	–	0.6	0.6	0.8
	Normal	81	0.11	0.14	91	0.70	0.77 ^a	≤1			
	High	114	0.29	0.25	95	0.67	0.70 ^a	–			
PLT (10 ⁹ /L)	Low	23	1.14	4.91	55	2.10	3.83	–	5.5	5.0	4.5
	Normal	274	2.72	0.99	175	4.90	2.80	≤4		3.0	3.0
	High	756	8.58	1.14	336	9.47	2.82	≤4		–	3.0
Neut (10 ⁹ /L)	Low	1.06	0.02	1.89	2.37	0.06	2.53	–	9.4	2.5	2.5
	Normal	3.62	0.08	2.21	3.84	0.09	2.34			2.5	
	High	19.73	0.18	0.91	13.23	0.22	1.66			–	
Lymph (10 ⁹ /L)	Low	0.49	0.01	2.04	0.58	0.02	3.45	–	7.9	3.5	3.5
	Normal	2.10	0.05	2.38	1.73	0.04	2.31				
	High	1.13	0.03	2.65	3.54	0.09	2.54				
Mono (10 ⁹ /L)	Low	0.01	0	0	0.09	0.01	11.11 ^{a,b}	–	10.5	8.5	8.5
	Normal	0.34	0.06	17.65 ^{a,b}	0.23	0.02	8.70 ^b				
	High	1.01	0.08	7.92	0.42	0.04	9.52 ^b				
Eo (10 ⁹ /L)	Low	0.01	0	0	0.86	0.05	5.81	–	11.3	10	10
	Normal	0.11	0.01	9.09	1.20	0.06	5.00				
	High	3.00	0.05	1.67	3.11	0.18	5.79				
Baso (10 ⁹ /L)	Low	0	0	–	0.05	0.01	20.00 ^a	–	9.5	20	20
	Normal	0.03	0.01	33.33 ^{a,b}	0.11	0.02	18.18 ^a				
	High	0.04	0.01	25.0 ^{a,b}	0.48	0.03	6.25				

^a: The CV values exceeded the minimum biological variation specifications; ^b: The CV values exceeded the state of art limits. MBV: Minimal biological variation; SOTA: State of art; WBC: White blood cell; RBC: red Blood cell; HGB: Hemoglobin; MCV: Mean corpuscular volume; PLT: Platelet; Neut: Neutrophile count; Lymph: Lymphocyte count; Mono: Monocyte count; Eo: Eosinophile count; Baso: Basophile count.

120–1143×10⁹/L. In addition, we assessed WBC linearity in LW mode over the concentration range of 0–1.83×10⁹/L. The differences between the measured and assigned values were within both the manufacturer's allowable deviation limits and the MBV TEa criteria. The linear regression results are presented in Table 3.

Comparison study

The method comparison results between the Maccura F810 and the Mindray BC-6200 are summarized in Table 4. According to the Passing-Bablok regression analysis, the 95% CIs for both the slope and intercept included 1.0 and 0.0, respectively, for MCV, lymphocyte count, and IG count. For HGB, hematocrit, PLT, and basophil count, the 95% CI for the slope included 1.0,

Table 2. Carry-over rates from high and low level samples

Parameters	Carry-over rate (%)	Maccura (%)	SOTA (%)
WBC	0.01	≤1	<0.2
RBC	0.30	≤1	<0.5
HGB	0	≤1	<0.4
PLT	0.46	≤1	<0.5

SOTA: State of the art; WBC: White blood cell; RBC: Red blood cell; HGB: Hemoglobin; PLT: Platelet.

whereas the 95% CI for the intercept did not include 0.0, showing a constant bias. In contrast, the intercept CIs for neutrophil and eosinophil counts included 0.0, while the slope CIs did not include 1.0, indicating a proportional bias. Neither the slope

Table 3. Linearity for Maccura F810

Parameter	Linear range	Regression	r ²	Maccura's allowable deviation range	MBV (TEa)
WBC (Low-level)	0–1.00×10 ⁹ /L	y=-0.014+1.013x	0.9989	0.20×10 ⁹ /L	21.3%
WBC (Mid-level)	0–25.45×10 ⁹ /L	y=0.106+0.989x	0.9992	2%	
WBC (High-level)	14.08–140.79×10 ⁹ /L	y=0.391+0.994x	0.9988	6%	
WBC (LW mode)	0–1.83×10 ⁹ /L	y=-0.001+0.987x	0.9984	0.20×10 ⁹ /L	
RBC	0–7.85×10 ¹² /L	y=0.077+1.003x	0.9986	2%	6%
HGB	0–20.83 g/dL	y=0.054+ 0.999x	0.9986	2%	5.8%
PLT (Low-level)	0–24.67×10 ⁹ /L	y=0.224+0.982x	0.9968	10×10 ⁹ /L	16.5%
PLT (High-level)	114–1143×10 ⁹ /L	y=9.620+0.995x	0.9991	6%	

MBV: Minimal biological variation; Tea: Total allowable error; WBC: White blood cell; RBC: Red blood cell; HGB: Hemoglobin; PLT: Platelet.

Table 4. Comparison of results between the Maccura F810 and the Mindray BC-6200

Parameters	Intercept (95% CI)	Slope (95% CI)	r	p	MD (95% CI) (%)	Maccura (%)	MBV (%)	SOTA (%)
WBC (10 ⁹ /L)	0.104 (0.057–0.141)	0.982 (0.976–0.988)	0.997	<0.0001	-0.50 (-0.93– -0.17)	15	7.6	4.4
RBC (10 ¹² /L)	0.096 (0.062–0.128)	0.973 (0.966–0.981)	0.997	<0.0001	-0.44 (-0.64– -0.26)	6	2.8	3.2
HGB (g/dL)	0.100 (0.100–0.100)	1.000 (1.000–1.000)	0.998	<0.0001	1.19 (0.96–1.39)	6	2.8	1.3
HCT (%)	0.741 (0.400–1.288)	0.991 (0.976–1.000)	0.987	<0.0001	1.02 (0.62–1.53)	9	2.3	1.8
MCV (fL)	0.279 (-1.879–2.175)	1.011 (0.988–1.035)	0.963	<0.0001	1.42 (0.93–1.79)	7	1.5	2
PLT (10 ⁹ /L)	4.000 (0.872–6.948)	1.000 (0.990–1.014)	0.990	<0.0001	1.80 (1.17–2.38)	20	7.5	6.4
Neut (10 ⁹ /L)	-0.016 (-0.043–0.011)	0.980 (0.974–0.986)	0.997	<0.0001	-2.40 (-2.80– -2.03)	NS	10.8	3.2
Lymp (10 ⁹ /L)	0.020 (-0.003–0.031)	1.000 (0.993–1.010)	0.995	<0.0001	0.87 (0.19–1.49)	NS	9.4	5
Mono (10 ⁹ /L)	0.017 (0.001–0.033)	1.098 (1.065–1.131)	0.967	<0.0001	12.77 (11.55–14.28)	NS	10.1	15
Eo (10 ⁹ /L)	-0.004 (-0.008–0.000)	0.931 (0.909–0.958)	0.896	<0.0001	-10.00 (-11.44– -8.00)	NS	25.9	13
Baso (10 ⁹ /L)	0.010 (0.010–0.010)	1.000 (1.000–1.000)	0.615	<0.0001	22.22 (0.00–28.57)	NS	11.7	32
IG (10 ⁹ /L)	0.000 (0.000–0.001)	1.000 (0.977–1.000)	0.718	<0.0001	0.00 (0.00–0.00)	NS	–	–

MBV: Minimal biological variation; SOTA: State of art; CI: Confidence interval; NS: Not specified; WBC: White blood cell; RBC: Red blood cell; HGB: Hemoglobin; HCT: Hematocrit; MCV: Mean corpuscular volume; PLT: Platelet; Neut: Neutrophile count; Lymp: Lymphocyte count; Mono: Monocyte count; Eo: Eosinophile count; Baso: Basophile count; IG: Immature granulocyte.

nor the intercept met the acceptance criteria for WBC, RBC, or monocyte count. The MD% obtained from the Bland-Altman analysis was within the SOTA limits and the manufacturer's allowable deviation ranges for all parameters. All parameters, except monocyte count, also met the MBV criteria (Fig. 1). For WBC, neutrophil count, HGB, and MCV, the predicted bias at all predefined medical decision points remained within both the MBV bias and SOTA limits. For PLT, the predicted bias at the medical decision points of 150 and 450×10⁹/L was within the MBV bias and SOTA limits, whereas the predicted bias at 50×10⁹/L exceeded both acceptance criteria (Table 5). The distribution of samples according to the predefined medical decision levels is presented in the Appendix 1.

Discussion

In response to the growing patient population and the increasing demand for laboratory testing, clinical laboratories must provide accurate, reliable, and timely blood cell counts and differential results. To meet these demands, manufacturers continue to develop hematology analyzers with improved throughput and the ability to measure an expanding range

of parameters. Before implementation in routine clinical practice, the analytical performance of these systems should be thoroughly verified by laboratory professionals. In the present study, we evaluated the analytical performance of the Maccura F810 hematology analyzer and compared it with that of the Mindray BC-6200, which is routinely used in our laboratory.

The Maccura F810 demonstrated good precision in both the within-run and between-batch studies. Within-run precision was assessed using normal, abnormal, low-, and high-level patient samples, whereas between-batch precision was evaluated using low-, normal-, and high-level quality control materials. All results met the manufacturer's performance specifications, and almost all parameters satisfied both the MBV and SOTA acceptance criteria for within-run and between-batch precision. The between-batch imprecision for MCV exceeded the MBV criterion but remained within the SOTA limit. Similar between-batch precision for MCV has been reported for other hematology analyzers [14, 21, 22]. This finding may be attributed to the limited long-term stability of the quality control materials, which is likely to have a greater impact on between-batch precision

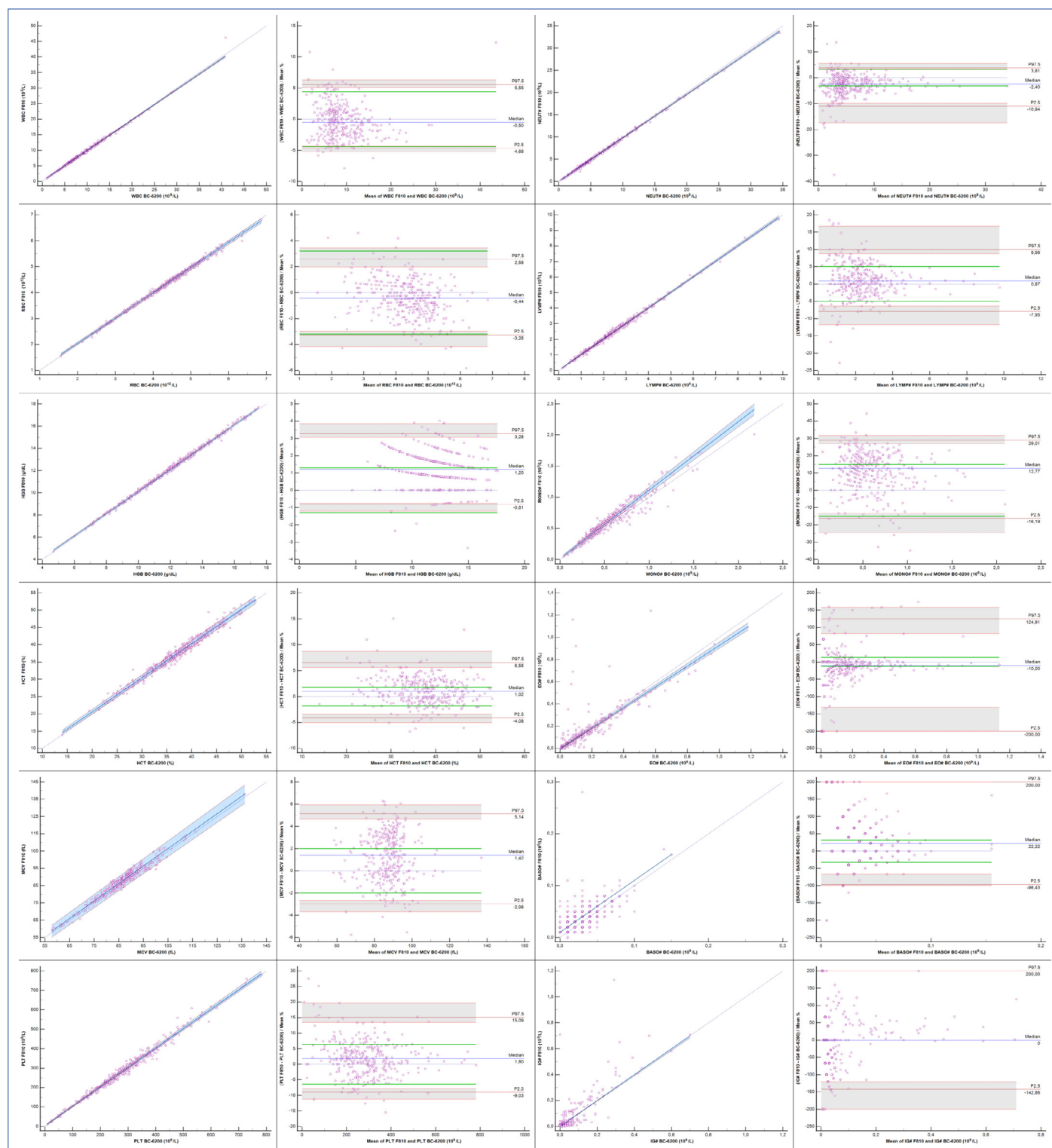


Figure 1. Comparison of CBC parameters between the Maccura F810 and Mindray BC-6200 hematology analyzers using Passing-Bablok regression and Bland-Altman analysis.

CBC: Complete blood count.

than on within-run precision. We evaluated the precision of monocyte, eosinophil, and basophil counts at very low concentrations. At such low concentrations, even minimal changes in the standard deviation can result in substantial

increases in the coefficient of variation. Consequently, relatively high CVs were observed for these parameters. A similar phenomenon has been described in previous analytical performance studies [4, 23, 24].

Table 5. Predicted bias at medical decision levels

Parameters	Medical decision level	Predicted result (95% CI)	Difference	Predicted bias (95% CI)	MBV (%)	SOTA (%)
WBC (10 ⁹ /L)	2	2.06 (2.02–2.09)	0.07 (0.03–0.09)	3.40 (1.43–4.80)	7.6	4.4
	4	4.03 (3.99–4.05)	0.03 (0.00–0.05)	0.79 (-0.09–1.33)		
	13	12.86 (12.81–12.90)	-0.13 (-0.18– -0.09)	-1.02 (-1.41– -0.72)		
Neut (10 ⁹ /L)	1.8	1.74 (1.7–1.76)	-0.05 (-0.07– -0.03)	-2.85 (-4.03– -1.93)	10.8	3.2
HGB (g/dL)	8	8.10 (8.10–8.20)	0.10 (0.10–0.20)	1.25 (1.25–2.50)	2.8	1.3
	12	12.10 (12.10–12.20)	0.10 (0.10–0.20)	0.83 (0.83–1.67)		
	16.5	16.60 (16.60–16.70)	0.10 (0.10–0.20)	0.61 (0.61–1.21)		
MCV (fL)	80	81.14 (80.73–81.50)	1.14 (0.73–1.50)	1.42 (0.91–1.87)	1.5	2
	100	101.35 (100.81–101.84)	1.35 (0.81–1.84)	1.35 (0.81–1.85)		
	150	154.00 (152.63–155.52)	4.00 (2.63–5.52)	2.67 (1.76–3.68)		
PLT (10 ⁹ /L)	50	54.00 (51.60–56.45)	4.00 (1.60–6.45)	8.00 (3.21–12.89)	7.5	6.4
	450	454.00 (451.32–457.91)	4.00 (1.32–7.91)	0.89 (0.29–1.76)		

MBV: Minimal biological variation; SOTA: State of art; CI: Confidence intervals; WBC: White blood cell; Neut: Neutrophile count; HGB: Hemoglobin; MCV: Mean corpuscular volume; PLT: Platelet.

In the routine workflow of clinical laboratories, samples with high cellular concentrations, such as those from patients with myeloproliferative neoplasms and lymphoproliferative disorders, are analyzed alongside cytopenic samples. To prevent interference and maintain analytical accuracy, it is necessary to minimize carry-over from high-concentration specimens into subsequently analyzed low-concentration specimens [8]. In the present study, the Maccura F810 demonstrated excellent carry-over performance, with all calculated carry-over rates remaining within both the manufacturer's specifications and the SOTA acceptance limits.

The widest concentration ranges attainable under routine laboratory conditions were selected to verify the manufacturer's linearity claims. No significant bias was seen across the evaluated concentration ranges, confirming excellent linearity. Because very low WBC and PLT levels may be encountered in routine clinical practice, we evaluated these ranges separately and also demonstrated excellent linearity. However, we could not verify the highest WBC (350–500×10⁹/L) and PLT (1001–5000×10⁹/L) concentrations claimed by the manufacturer because neither suitable patient samples nor commercially available linearity materials spanning these ranges were available at the time of the study.

In the method comparison study, MCV and lymphocyte count showed excellent agreement between the two analyzers. For the remaining parameters, either constant or proportional bias was identified; however, the calculated MD% remained within the SOTA acceptance limits for all parameters, indicating good agreement between the two systems. In addition, all parameters except monocyte count met the MBV acceptance criteria. Overall, the results confirmed the manufacturer's performance specifications for method comparison.

Monocytes are morphologically heterogeneous cells, and their accurate classification largely depends on the detection technology used by hematology analyzers. The compar-

ative study included samples from intensive care units and inpatient wards. In patients with sepsis or other infections, monocytes exhibit considerable morphological variability, particularly during the early stages of disease [25, 26]. In addition to methodological discrepancies between analyzers, this morphological heterogeneity is likely to contribute to both constant and proportional bias. In the present study, the 95% CIs for both the intercept and slope in the Passing-Bablok regression analysis of monocyte count excluded 0.0 and 1.0, respectively. Furthermore, the MD% exceeded the allowable MBV limit. These outcomes are consistent with those reported by İncir and Palaoğlu [11], who also observed a large mean difference in monocyte counts, which they attributed to the morphological complexity of these cells. Similar differences in monocyte counts have also been reported in studies comparing hematology analyzers or automated methods with manual differential counting [14, 27].

For basophil measurements, the median percentage difference between the two methods was 22.2% (95% CI, 0.0%–28.6%). Although the upper limit of the CI exceeded the MBV acceptance limit, the relatively large percentage difference is more likely attributable to the very low absolute basophil counts than to a systematic bias between the two methods. As basophil counts approach zero, even minimal absolute differences can produce disproportionately large relative differences. In addition, differences in measurement principles may contribute to variability among hematology analyzers [24]. The relatively lower correlation coefficients for basophil count ($r=0.615$) and IG ($r=0.718$) are primarily related to the narrow concentration ranges and clustering of results at low concentrations, as shown in the Passing-Bablok regression plots (Fig. 1). No specific acceptance specification has been established for IG. However, Passing-Bablok regression yielded a slope of 1.000 (95% CI, 0.977–1.000) and an intercept of 0.000 (95% CI, 0.000–0.001), indicating no significant proportional or systematic bias between the methods. A limitation of this study

is the low number of samples with a prominent left shift and markedly elevated IG values, which restricts performance assessment at higher concentration ranges.

In the present study, MD% for PLT obtained from the Bland-Altman analysis was 1.8% (95% CI, 1.17%-2.38%), remaining within both the SOTA and MBV bias limits. However, at the predefined medical decision level of $50 \times 10^9/L$, the predicted bias was 8.0%, exceeding both the MBV (7.5%) and SOTA (6.4%) acceptance limits, although its 95% CI (3.21%-12.89%) included both allowable limits. This estimate was associated with a relatively wide 95% CI, most likely because of the limited number of samples with severe thrombocytopenia included in the study (Appendix 1). On the other hand, the absolute difference at this decision level was $4 \times 10^9/L$, corresponding to a relatively small change in PLT concentration, and the bias may have limited clinical significance. Given the importance of PLT in transfusion and perioperative decision-making, laboratories introducing or replacing hematology analyzers should conduct local verification studies to determine whether PLT measurements among different analytical systems are interchangeable, particularly for thrombocytopenic samples.

Limitations

This study has several limitations. Because our hospital does not have a hematology department, extremely low and high cell-count samples were encountered relatively infrequently in routine practice, resulting in a limited number of specimens at the extremes of the analytical measuring range. Another important limitation was that the diagnostic performance of analyzer flags compared with peripheral blood smear review could not be evaluated because an adequate number of qualified examiners was unavailable. In addition, although the Maccura F810 is capable of reticulocyte analysis, we were unable to assess this feature due to logistical and technical issues. Finally, the analytical performance of the analyzer could not be compared with that of reference methods.

Conclusion

In the present study, we evaluated the performance of the Maccura F810 hematology analyzer against the manufacturer's specifications and confirmed all performance claims. Because the manufacturer's acceptance criteria are relatively broad, the analytical performance was also assessed using the more stringent SOTA and MBV criteria to better reflect the requirements of our laboratory [3, 13]. To the best of our knowledge, this is the first verification study of the Maccura F810 hematology analyzer reported in the literature. Overall, the Maccura F810 demonstrated satisfactory analytical performance in terms of precision, carry-over, linearity, and method comparison. However, the predicted PLT bias at the $50 \times 10^9/L$ medical decision level exceeded both the MBV and SOTA acceptance limits, highlighting the need for careful local verification of PLT results in thrombocytopenic samples. The results indicate that the analyzer is well suited to meet the demands of high-throughput clinical laboratories..

Disclosures

Ethics Committee Approval: The study was approved by the Çanakkale Onsekiz Mart University Clinical Research Ethics Committee (no: 2023/01-18, date: 04/01/2023).

Informed Consent: The study used anonymized residual routine laboratory samples; informed consent was not applicable.

Conflict of Interest Statement: İz Tıbbi Cihaz Medikal San. ve Tic. Ltd. Şti. (Türkiye), the Turkish distributor of the Maccura F810 hematology analyzer, provided the analyzer, reagents, and consumables used in this study. The company had no role in the study design, data collection, data analysis, interpretation of the results, manuscript preparation, or the decision to submit the manuscript for publication. The authors received no personal financial support or other compensation from the company.

Funding: This study received in-kind support from İz Tıbbi Cihaz Medikal San. ve Tic. Ltd. Şti. (Türkiye), which provided the Maccura F810 hematology analyzer, reagents, and consumables used in the study.

Use of AI for Writing Assistance: AI tools were used only for grammar checking and language editing during the preparation of this manuscript. The study design, data collection, data analysis, interpretation of the results, and conclusions were performed by the authors in accordance with scientific research methods and academic ethical principles. The authors take full responsibility for the content of the manuscript.

Authorship Contributions: Concept – H.Y.C., M.H.S.B.; Design – H.Y.C., M.H.S.B.; Supervision – H.Y.C., M.H.S.B.; Funding – H.Y.C.; Materials – H.Y.C.; Data collection and/or processing – H.Y.C.; Data analysis and/or interpretation – H.Y.C., M.H.S.B.; Literature search – H.Y.C., M.H.S.B.; Writing – H.Y.C., M.H.S.B.; Critical review – H.Y.C., M.H.S.B.

Acknowledgments: The authors sincerely thank the laboratory technicians of the Çanakkale Onsekiz Mart University Hospital Medical Biochemistry Laboratory for their valuable technical assistance and support during the study.

Peer-review: Externally peer-reviewed.

References

1. Verbrugge SE, Huisman A. Verification and standardization of blood cell counters for routine clinical laboratory tests. *Clin Lab Med* 2015;35(1):183–96. [CrossRef]
2. Vajpayee N, Graham SS, Bem S. Basic examination of blood and bone marrow. In: McPherson RA, Pincus MR, editors. *Henry's clinical diagnosis and management by laboratory methods*. 23rd ed. St. Louis (MO): Elsevier; 2017. p. 510–39.
3. Vis JY, Huisman A. Verification and quality control of routine hematology analyzers. *Int J Lab Hematol* 2016;38(Suppl 1):100–9. [CrossRef]
4. Seyithanoğlu M, Tuncer Hİ, Tolun Fİ, Kiliç M. Comparison of results of two hematological analyzer systems: Dirui BF-7200 and Sysmex XN-1000. *Türk Biyokimya Dergisi* 2024;49(2):157–66. [CrossRef]

5. Yucel D, Serdar MA. A brief history of hematology analyzers and recent advancements: the available testing wealth. *Türk Biyokimya Dergisi* 2024;49(3):303–5. [\[CrossRef\]](#)
6. Daves M, Roccaforte V, Lombardi F, Panella R, Pastori S, Spreafico M, et al. Modern hematology analyzers: beyond the simple blood cells count (with focus on the red blood cells). *J Lab Precis Med* 2024;9:4. [\[CrossRef\]](#)
7. Hotton J, Broothaers J, Swaelens C, Cantinieaux B. Performance and abnormal cell flagging comparisons of three automated blood cell counters: Cell-Dyn Sapphire, DxH-800, and XN-2000. *Am J Clin Pathol* 2013;140(6):845–52. [\[CrossRef\]](#)
8. Briggs C, Culp N, Davis B, d'Onofrio G, Zini G, Machin SJ. ICSH guidelines for the evaluation of blood cell analysers including those used for differential leucocyte and reticulocyte counting. *Int J Lab Hematol* 2014;36(6):613–27. [\[CrossRef\]](#)
9. Clinical and Laboratory Standards Institute (CLSI). Validation, verification, and quality assurance of automated hematology analyzers. 2nd ed. Wayne (PA): Clinical and Laboratory Standards Institute; 2010. CLSI document H26–A2.
10. International Organization for Standardization. ISO 15189:2022 Medical laboratories-Requirements for quality and competence. Geneva: International Organization for Standardization; 2022. Available at: <https://www.iso.org/standard/76677.html>. Accessed: June 9, 2026.
11. Incir S, Palaoglu KE. Evaluation of the performance of sysmex XN-3100 automated hematology analyzer regarding the sysmex XE-2100 and microscopic examination. *Turkish Journal of Biochemistry* 2021;46(1):29–37. [\[CrossRef\]](#)
12. Bayraktar N, Topcu DI. User verification of Abbott Alinity HQ hematology analyzer. *Turkish Journal of Biochemistry* 2024;49(2):167–77. [\[CrossRef\]](#)
13. Aarsand AK, Webster C, Fernandez-Calle P, Jonker N, Diaz-Garzon J, Coskun A, et al. The EFLM Biological Variation Database. Available at: <https://biologicalvariation.eu/> Accessed: 10 June, 2026.
14. Kulik K, Kwiecień I, Chełstowska B, Rutkowska E, Rzepecki P. Evaluation and comparison of the new Mindray BC-6200 hematology analyzer with ADVIA 2120i. *Int J Lab Hematol* 2021;43(3):395–402. [\[CrossRef\]](#)
15. Clinical and Laboratory Standards Institute. Measurement procedure comparison and bias estimation using patient samples. 3rd ed. Wayne (PA): Clinical and Laboratory Standards Institute; 2018. CLSI document EP09c.
16. Fioredda F, Skokowa J, Tamary H, Spanoudakis M, Farruggia P, Almeida A, et al. The European Guidelines on Diagnosis and Management of Neutropenia in Adults and Children: A Consensus Between the European Hematology Association and the EuNet-INNOCHRON COST Action. *Hemasphere* 2023;7(4):e872. [\[CrossRef\]](#)
17. Khoury JD, Solary E, Abia O, Akkari Y, Alaggio R, Apperley JF, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* 2022;36(7):1703–19. [\[CrossRef\]](#)
18. Arber DA, Orazi A, Hasserjian RP, Borowitz MJ, Calvo KR, Kvasnicka HM, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* 2022;140(11):1200–28. [\[CrossRef\]](#)
19. World Health Organization. Guideline on haemoglobin cutoffs to define anaemia in individuals and populations. Available at: <https://www.who.int/publications/i/item/9789240088542> Accessed 15 June, 2026.
20. General Directorate of Health Services Department of Examination And Diagnosis Services. Decision Limit (Threshold Value), Critical Value (Panic Value) and Harmonization Procedure of Measurement Units within the framework of the “Rational Laboratory Use Project”. Available at: <https://shgmtetikdb.saglik.gov.tr/TR-32805/akilci-laboratuvar-kullanimi-projesi-cercevesinde-karar-siniri-esik-deger-kritik-deger-panik-deger-ve-olcum-birimlerinin-harmonizasyonu-proseduru.html> Accessed 15 June, 2026.
21. Woo KS, Jeong IH, An GD, Lim HH, Han JY. Performance evaluation of new Abbott Alinity hq hematology analyzer. *Int J Lab Hematol* 2019;41(5):657–63. [\[CrossRef\]](#)
22. Van der Beken Y, Van Dalem A, Van Moer G, Segers E, Damiens S, Hoffmann J, et al. Performance evaluation of the prototype Abbott Alinity hq hematology analyzer. *Int J Lab Hematol* 2019;41(4):448–55. [\[CrossRef\]](#)
23. Lee H, Cho J, Yoo G, Kim J. Analytical and Clinical Performance of the Atellica HEMA 580 Haematology Analyzer: A Comparison With the ADVIA 2120i. *Int J Lab Hematol* 2026;48(4):784–94. [\[CrossRef\]](#)
24. Slim CL, Wevers BA, Demmers M, Lakos G, Hoffmann J, Adriaansen HJ, et al. Multicenter performance evaluation of the Abbott Alinity hq hematology analyzer. *Clin Chem Lab Med* 2019;57(12):1988–98. [\[CrossRef\]](#)
25. Agnello L, Vidali M, Lo Sasso B, Giglio RV, Gambino CM, Scazzone C, et al. Monocyte distribution width (MDW) as a screening tool for early detecting sepsis: a systematic review and meta-analysis. *Clin Chem Lab Med* 2022;60(5):786–92. [\[CrossRef\]](#)
26. Piva E, Zuin J, Pelloso M, Tosato F, Fogar P, Plebani M. Monocyte distribution width (MDW) parameter as a sepsis indicator in intensive care units. *Clin Chem Lab Med* 2021;59(7):1307–14. [\[CrossRef\]](#)
27. Velizarova M, Yacheva T, Genova M, Svinarov D. Evaluation of automated hematology analyzer DYMIND DH76 compared to SYSMEX XN 1000 system. *J Med Biochem* 2021;40(4):367–77. [\[CrossRef\]](#)

Appendix 1. Distribution of samples across medical decision levels

Parameter	Interval	n (%)
WBC ($10^9/L$)	≤ 2.00	5 (1.4)
	2.01–4.00	27 (7.6)
	4.01–13.00	271 (76.1)
	> 13.00	53 (14.9)
Neut ($10^9/L$)	≤ 1.80	25 (7.0)
	> 1.80	331 (93.0)
HGB (g/dL)	≤ 8.0	20 (5.6)
	8.1–12.0	151 (42.4)
	12.1–16.5	175 (49.2)
	> 16.5	10 (2.8)
MCV (fL)	≤ 80	70 (19.7)
	81–100	263 (73.9)
	> 100	23 (6.5)
PLT ($10^9/L$)	≤ 50	8 (2.2)
	51–150	31 (8.7)
	151–450	287 (80.6)
	> 450	30 (8.4)

The total number of samples was 356 for each parameter. WBC: White blood cell; Neut: Neutrophile count; HGB: Hemoglobin; MCV: Mean corpuscular volume; PLT: Platelet.



Research Article

Interpretation of heterophile-blocking tube results in thyroid function testing using analytical change limits: An exploratory study of clinically discordant cases

Rana Turkal¹, Ali Yaman¹, Tulay Cevlik¹, Hulya Gozu², Onder Sirikci^{1,3}, Goncagul Haklar^{1,3}

¹Biochemistry Laboratory, Marmara University Pendik Training and Research Hospital, Istanbul, Türkiye

²Division of Endocrinology, Department of Internal Medicine, Marmara University Faculty of Medicine, Istanbul, Türkiye

³Department of Biochemistry, Marmara University Faculty of Medicine, Istanbul, Türkiye

Abstract

Objectives: This study aimed to evaluate an analytical framework for interpreting heterophile-blocking tube (HBT) results in thyroid function tests (TFTs) by comparing HBT-associated changes in thyroid-stimulating hormone (TSH), free thyroxine (fT4), and free triiodothyronine (fT3) with analyte-specific analytical change limits (ACLs), documenting transitions across laboratory reference intervals, and comparing changes across the three analytes measured in the same treated aliquot.

Methods: This single-center, prospective, exploratory observational study enrolled 41 adults whose TFT results were discordant with their clinical presentations. Untreated and HBT-treated aliquots of the same serum specimen were analyzed in the same run. Signed percentage changes were compared with ACLs of $\pm 2.77 \times$ the analytical coefficient of variation (CV_a) (12.47% for TSH, 15.84% for fT3, and 14.40% for fT4), and paired concentrations were compared using the Wilcoxon signed-rank test with Holm adjustment. Changes in the three analytes from the same treated aliquot were compared, and the sensitivity of the classification to the imprecision estimate was examined.

Results: Median signed changes were -0.45% for TSH, -6.47% for fT3, and -2.11% for fT4; only fT3 showed a significant paired reduction (adjusted $p=0.0038$). The ACL was exceeded in 8/41 (19.5%) fT3, 1/39 (2.6%) TSH, and 2/41 (4.9%) fT4 results. The rates of TSH and fT4 exceedance were close to those expected from analytical variation alone ($p=0.86$ and $p=0.61$, respectively), whereas fT3 exceeded the limit approximately four times more often ($p=0.0008$). fT3 changes showed no significant correlation with concurrent changes in TSH or fT4, clustered in specimens with elevated pre-HBT fT3 levels, and remained more frequent at every assumed CV_a ranging from 1.8% to 8%. Threshold-supported normalization occurred in four fT3 results and in none of the TSH or fT4 results.

Conclusion: An imprecision-based ACL combined with reference-interval transition provides a transparent criterion for HBT interpretation. HBT-responsive changes occurred predominantly in fT3, whereas TSH and fT4 in the same aliquots were largely unaffected, arguing against a nonspecific pretreatment effect. A change exceeding the ACL is compatible with, but does not by itself prove, heterophile antibody interference and should be interpreted together with the complete thyroid profile, the clinical context, and, where necessary, independent confirmatory testing.

Keywords: Endocrinology, heterophile antibodies, immunoassay, thyroid function tests

How to cite this article: Turkal R, Yaman A, Cevlik T, Gozu H, Sirikci O, Haklar G. Interpretation of heterophile-blocking tube results in thyroid function testing using analytical change limits: An exploratory study of clinically discordant cases. Int J Med Biochem 2026;9(4):232–241.

Automated immunoassays are the methods of choice for thyroid function testing, with thyroid-stimulating hormone (TSH) and free thyroxine (fT4) as the primary paramet-

ters and free triiodothyronine (fT3) as a complementary test. Nevertheless, thyroid immunoassays remain susceptible to analytical interference from macro-TSH, biotin, anti-strepta-

Address for correspondence: Rana Turkal, MD. Biochemistry Laboratory, Marmara University Pendik Training and Research Hospital, Istanbul, Türkiye

Phone: +90 216 625 45 45 **E-mail:** ranaturkal@yahoo.com **ORCID:** 0000-0002-0042-7261

Submitted: August 20, 2026 **Revised:** September 22, 2026 **Accepted:** September 24, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



vidin and anti-ruthenium antibodies, thyroid hormone autoantibodies, heterophile antibodies, and rheumatoid factor [1–4]. The 2026 European Thyroid Association (ETA) guideline emphasizes that susceptibility to interference is assay-specific and recommends that endocrinologists and laboratory specialists be familiar with the assay used for a particular patient and its limitations [5, 6].

Heterophile antibodies are endogenous immunoglobulins with low-affinity, polyspecific binding to animal immunoglobulins, including assay reagent antibodies; by bridging or blocking capture and detection antibodies, they may produce falsely increased or decreased results. They are distinguished from specific human anti-animal antibodies that develop after exposure to animal-derived immunoglobulins, although immunization, transfusion, autoimmune disease, monoclonal antibody therapy, animal contact, or infection may raise suspicion [1, 2, 7, 8]. Because interference depends on assay architecture, reagent antibodies, and blocking systems, the same sample may yield markedly different results across different platforms, and interference cannot be predicted from clinical findings alone [2, 9, 10]. Reported frequencies range from approximately 0.05% to 6%, depending on the assay, analyte, population, and criteria used to define interference; unrecognized interference in thyroid testing may prompt unnecessary investigations and inappropriate treatment [2, 7, 9].

Suspected interference can be investigated by remeasurement using an alternative assay, serial dilution, polyethylene glycol (PEG) precipitation, or treatment with a heterophile-blocking reagent [11–13]. Heterophile-blocking tubes (HBTs) are a practical option for routine laboratories, but the response to blocking varies according to the interfering antibody, blocking reagent, and assay design; an unchanged result does not exclude interference, and a change does not identify the interfering substance [1, 11]. The 2026 ETA guideline lists HBT pretreatment among the recommended investigative tools, describing it as effective for heterophilic antibody interference but limited to known heterophilic interactions and cautioning that results may be assay-dependent and that false positives may occur; however, it does not specify how large a post-HBT change must be to be regarded as meaningful [5]. Responses to blocking reagents are assay-specific, and HBT-related artifacts have been described for certain assay designs [14–16]. Objective analytical criteria for interpreting HBT results therefore remain limited.

Recognition of interference depends on communication between clinicians and laboratory specialists when the thyroid function test (TFT) pattern is internally inconsistent or does not fit the clinical picture [5, 13, 17, 18]. Published blocking studies have used assay-imprecision-based or empirically derived thresholds [19, 20] or group-level paired tests alone [21]. We therefore evaluated an analytical framework for interpreting HBT results in clinically selected patients with discordant TFTs by comparing changes in TSH, fT4, and fT3 after HBT treatment with analyte-specific analytical change limits and documenting reference-interval transitions.

Materials and Methods

Study population

This single-center, prospective, exploratory observational study was conducted over a 10-month period, during this period, 17,989 TFT panels were performed for 8,299 unique patients evaluated at the Endocrinology Outpatient Clinic of Marmara University Pendik Training and Research Hospital. Forty-one adults were selected by endocrinologists for HBT evaluation because their TFT results were discordant with their clinical presentation; HBT testing was requested on a case-by-case basis, and patients were not systematically screened or consecutively enrolled. Eligibility required at least one of the following: an internally inconsistent TFT profile, biochemical findings that did not explain the clinical presentation, absence of the expected biochemical response to treatment, or hormone results incompatible with the suspected thyroid disorder. Case-based selection reflects the setting in which HBTs are used in practice, as current guidance recommends considering assay interference when TFTs are incongruent with the clinical picture or with each other [5]. This study was approved by the Marmara University Faculty of Medicine Ethics Committee (Date: 05.09.2025, Decision No.: 09.2025.25-0760), and all participants provided informed consent.

Clinical questionnaire

A standardized questionnaire administered at the Biochemistry Laboratory documented demographic characteristics, occupation, transfusion history, allergic disease, dietary habits, pet ownership or other animal exposure, chronic diseases (including autoimmune, thyroid, rheumatological, and cardiovascular diseases), reason for admission, recent infections, vaccinations, medication use, dialysis, and previous therapies, in line with recommendations for clinical review in patients with suspected assay interference [2, 5]. The questionnaire was descriptive; the study was not designed to test associations between exposures and HBT responses.

Sample collection and preliminary assessment

Venous blood was collected into gel-containing serum separator tubes, allowed to clot for 30 minutes, and centrifuged at approximately 2190×g for 10 minutes. Each serum specimen was visually inspected, and lipemia, icterus, and hemolysis (LIH) indices were measured using a Beckman Coulter AU5800 analyzer (Beckman Coulter, Brea, CA, USA); all 41 specimens were within the normal category for all three indices, and none were excluded. Before HBT testing, calibration status, internal quality-control results, reagent status, analyzer flags, and sample identification were reviewed; internal quality control was within acceptance limits, and satisfactory performance for TSH, fT3, and fT4 was confirmed through the Randox International Quality Assessment Scheme (RIQAS). HBT testing was undertaken only when no preanalytical, analytical, or postanalytical error explained the discordance.

Heterophile-blocking procedure and thyroid hormone measurements

Heterophilic Blocking Tubes (part No. 3IX762, Scantibodies Laboratory, Santee, CA, USA) contain a proprietary lyophilized reagent with binders intended to inactivate heterophile antibodies; because the reagent is lyophilized, the procedure adds no diluent. In accordance with the manufacturer's instructions, 500 μ L of serum was transferred into an HBT, mixed by five inversions, and incubated for 60 minutes at 18–28°C. Paired untreated and HBT-treated aliquots from the same specimen were analyzed in the same run using a UniCel Dxl 800 analyzer (Beckman Coulter). TSH was measured using a two-site sandwich chemiluminescent immunoassay (Access HYPERSensitive hTSH), whereas fT3 and fT4 were measured using competitive chemiluminescent immunoassays (Access Free T3 and Access Free T4) [22]. HBT-treated results were used solely for within-pair comparisons and were not reported as patient results.

Because HBT-related over-recovery has been reported for Access assays whose solid phase consists of goat anti-mouse antibody-coated particles, attributable to murine components of the blocking reagent [15, 16], assay architecture was verified using the manufacturer's instructions for use [22]. The HYPERSensitive hTSH assay uses paramagnetic particles coated with goat anti-mouse immunoglobulin G (IgG) carrying mouse monoclonal anti-hTSH complexes and therefore has the design feature for which HBT-related over-recovery has been described. The Free T3 assay uses a monoclonal anti-T3-alkaline phosphatase conjugate with streptavidin-coated particles and a biotinylated T3 analog, whereas the Free T4 assay uses a biotinylated mouse monoclonal anti-T4 antibody with streptavidin-coated particles; neither free-hormone assay uses a goat anti-mouse solid phase. TSH results after HBT treatment were therefore examined specifically for over-recovery.

Laboratory reference intervals were 0.34–5.60 mIU/L for TSH, 2.60–4.37 ng/L for fT3, and 0.61–1.12 ng/dL for fT4. The TSH reportable range was 0.015–100 mIU/L (functional sensitivity, 0.015 mIU/L; lower values were reported as <0.015 mIU/L); the measuring intervals were approximately 0.88–30 ng/L for fT3 and 0.25–6.0 ng/dL for fT4.

Assessment of changes following HBT treatment

For each analyte, the effect of HBT treatment was expressed as the signed percentage change: $100 \times (\text{HBT-treated} - \text{untreated}) / \text{untreated}$. To identify changes exceeding analytical imprecision, an analyte-specific analytical change limit (ACL) was calculated at the two-sided 95% level as $ACL_{95} = 1.96 \times \sqrt{2} \times CV_a = 2.77 \times CV_a$, where CV_a is the analytical coefficient of variation and $\sqrt{2}$ accounts for the imprecision of two measurements [23, 24]. Because the paired measurements represented an in vitro comparison of aliquots from the same specimen, within-subject biological variation was not included, and a conventional reference change value was not applied.

Laboratory-derived between-day CV_a values (4.50% for TSH, 5.20% for fT4, and 5.72% for fT3; cumulative internal quality-control data from two control levels) were used, yielding ACLs of $\pm 12.47\%$, $\pm 14.40\%$, and $\pm 15.84\%$, respectively. Between-day imprecision comprises the within-run component together with between-run and between-day components and is therefore expected to be at least as large as within-run imprecision [23]; the resulting ACL may consequently have yielded relatively conservative thresholds for same-run comparisons, such that classification would tend toward undercalling rather than overcalling threshold exceedance, although imprecision varies with concentration and was not separately established for HBT-treated serum. The manufacturer's stated within-run CV_a values at control concentrations within and above the reference intervals are 2.5–3.1% for TSH, 1.8–2.7% for fT4, and 2.6–5.1% for fT3 and are higher only at the lowest control level of each assay [22]; the sensitivity analysis described below (1.8–8%) encompasses all of these values. An absolute change exceeding the ACL was classified as exceeding the change expected from analytical variation alone and is termed an HBT-responsive change because HBT treatment does not identify the interfering substance.

Exploratory cross-analyte and sensitivity analyses

Because TSH, fT3, and fT4 were measured in the same HBT-treated aliquot, a pretreatment effect common to the entire aliquot (volume displacement or gross matrix alteration) would be expected to affect all three analytes. Comparisons across analytes therefore address a nonspecific pretreatment effect and complement an HBT-treated control group. Three post hoc analyses were added during peer review and are reported as exploratory. First, observed ACL exceedance for each analyte was compared with the nominal 5% rate of a two-sided 95% limit using the one-sided exact binomial test; this nominal rate applies only if the CV_a used reflects the imprecision of the paired measurements under the study conditions. Exceedance was compared across analytes using Cochran's Q test for related proportions in the 39 patients with quantifiable pairs for all three analytes, with an exact p value obtained by permutation within patients. Second, Spearman correlation was used to assess whether percentage changes were related across analytes within specimens and whether the fT3 change was related to the pre-HBT fT3 concentration; fT3 changes were also summarized according to whether pre-HBT fT3 exceeded the upper reference limit, and the paired fT3 comparison was repeated after excluding specimens exceeding the ACL. Third, ACL classification was recalculated for assumed CV_a values of 1.8–8%.

Statistical analysis

Continuous variables were summarized as the median (interquartile range [IQR]). Paired pre- and post-HBT comparisons for the three analytes were the primary analyses and were performed using the two-sided asymptotic Wilcoxon signed-rank test, with Holm adjustment for the three comparisons; a

Table 1. Comparison of thyroid function test results before and after heterophile-blocking treatment

Analyte	Evaluable pairs, n	Pre-HBT, median (IQR)	Post-HBT, median (IQR)	Signed change, %, median (IQR)	Paired Wilcoxon P	Holm-adjusted P ^b	Analytical threshold exceeded, n/N (%) [95% CI]	Post-HBT normalization with ACL exceedance, n/N (%) [95% CI]
TSH (mIU/L)	39 ^a	2.36 (1.57–5.06)	2.34 (1.52–4.75)	-0.45 (-3.71 to +1.74)	0.2519	0.2519	1/39 (2.6%) [0.5–13.2%]	0/39 (0.0%) [0.0–9.0%]
ft3 (ng/L)	41	3.66 (3.28–5.00)	3.56 (3.15–3.96)	-6.47 (-13.91 to +1.08)	0.0013	0.0038	8/41 (19.5%) [10.2–34.0%]	4/41 (9.8%) [3.9–22.5%]
ft4 (ng/dL)	41	1.18 (0.95–1.55)	1.15 (0.91–1.53)	-2.11 (-5.77 to +2.59)	0.0389	0.0778	2/41 (4.9%) [1.3–16.1%]	0/41 (0.0%) [0.0–8.6%]

Data are presented as median (interquartile range). Signed percentage change was calculated as $100 \times (\text{post-HBT} - \text{pre-HBT}) / \text{pre-HBT}$. Paired P values were obtained using the two-sided asymptotic Wilcoxon signed-rank test and adjusted for the three analytes using the Holm method. Analytical change limits (ACLs) were defined as $\pm 2.77 \times \text{CV}_s$; 12.47% for TSH, 15.84% for ft3, and 14.40% for ft4. Post-HBT normalization with ACL exceedance required both movement from outside to within the corresponding reference interval (TSH, 0.34–5.60 mIU/L; ft3, 2.60–4.37 ng/L; ft4, 0.61–1.12 ng/dL) and an absolute percentage change exceeding the corresponding ACL. Wilson score 95% confidence intervals were calculated for the proportions reported in the final two columns. For n/N, n denotes the number of pairs meeting the specified criterion and N the number of evaluable pairs.

^a: TSH paired concentration and percentage-change analyses included 39 quantifiable pairs because two pairs were reported as <0.015 mIU/L both before and after HBT treatment;

^b: A two-sided Holm-adjusted $p < 0.05$ was considered statistically significant. ACL: Analytical change limit; CI: Confidence interval; CV_s : Analytical coefficient of variation; ft3: Free triiodothyronine; ft4: Free thyroxine; HBT: Heterophile-blocking tube; IQR: Interquartile range; TSH: Thyroid-stimulating hormone.

Holm-adjusted $p < 0.05$ was considered statistically significant. ACL exceedance and reference-interval transitions were descriptive secondary outcomes. Post-HBT normalization with ACL exceedance was defined as a transition from outside to within the reference interval accompanied by an absolute change exceeding the ACL; Wilson score 95% confidence intervals were calculated for proportions. TSH pairs reported as <0.015 mIU/L both before and after treatment were excluded from paired concentration and percentage-change analyses but retained for reference-interval transition analyses; zero differences were excluded from the signed-rank calculation. Analyses were performed using Python 3.12.13 with NumPy 2.3.5 and SciPy 1.17.0; figures were generated using Matplotlib 3.10.8.

Results

Study cohort

The clinical and exposure characteristics of the 41 patients are summarized in Appendix 1. The median age was 41 years (IQR, 31–59; range, 22–74), and 27 patients (65.9%) were female. Multiple-analyte abnormalities were the most frequent pre-HBT pattern (16 patients, 39.0%), followed by isolated ft4 abnormalities (10, 24.4%); all three analytes were within the reference intervals in 7 patients (17.1%), whereas isolated ft3 and isolated TSH abnormalities were observed in 5 (12.2%) and 3 (7.3%) patients, respectively. Known thyroid disease was recorded in 20/38 (52.6%), animal contact in 12/40 (30.0%), and recent upper respiratory tract infection or gastroenteritis in 7/40 (17.5%) of patients with available information.

Cohort-level paired comparisons and specimen-level ACL exceedance

Paired ft3 and ft4 results were available for all 41 patients; TSH percentage-change analyses included 39 pairs because both results for patients 30 and 34 were <0.015 mIU/L (Table 1). Median TSH was 2.363 mIU/L (IQR, 1.565–5.064) before and 2.343 mIU/L (IQR, 1.523–4.747) after treatment; the median

signed change was -0.45% (IQR, -3.71% to +1.74%; Holm-adjusted $p = 0.2519$). One of 39 pairs (2.6%) exceeded the TSH ACL (+12.66%), with both measurements within the reference interval. Median ft4 decreased from 1.18 ng/dL (IQR, 0.95–1.55) to 1.15 ng/dL (IQR, 0.91–1.53), representing a median change of -2.11% (IQR, -5.77% to +2.59%); the difference was significant before ($p = 0.0389$) but not after Holm adjustment ($p = 0.0778$). Two of 41 pairs (4.9%) exceeded the ft4 ACL, neither with normalization. Median ft3 decreased from 3.66 ng/L (IQR, 3.28–5.00) to 3.56 ng/L (IQR, 3.15–3.96), representing a median change of -6.47% (IQR, -13.91% to +1.08%), which was within the ft3 ACL; the reduction remained significant after adjustment ($p = 0.0013$; Holm-adjusted $p = 0.0038$). Eight of 41 pairs (19.5%) exceeded the ft3 ACL (patients 1, 3, 4, 22, 26, 32, 33, and 40; Fig. 1): seven decreases ranging from -16.3% to -38.6% and one increase of +35.6% (patient 22).

Reference-interval transitions

Individual ft3 trajectories are shown in Figure 2, and the eight patients with threshold-exceeding changes are described in Table 2. Four patients (patients 1, 3, 4, and 40) moved from above to within the ft3 reference interval; patients 26 and 32 remained within the interval, patient 33 remained above it, and patient 22 remained below it. Evaluated independently of the threshold, post-HBT normalization occurred in five ft3 results (12.2%; patients 1, 3, 4, 7, and 40) and two ft4 results (4.9%; patients 15 and 34), but in no TSH result; patient 7 reached the ft3 upper reference limit with a change (-5.6%) within the ACL. Thus, four ft3 results (9.8%) fulfilled both the normalization and threshold criteria (Table 1), whereas neither ft4 normalization exceeded the ACL, and two additional ft4 results moved from within to outside the interval.

Exploratory cross-analyte and sensitivity analyses

Under a two-sided 95% ACL, approximately 5% of pairs are expected to exceed the limit by chance, provided that the imprecision estimate reflects the paired measurements under

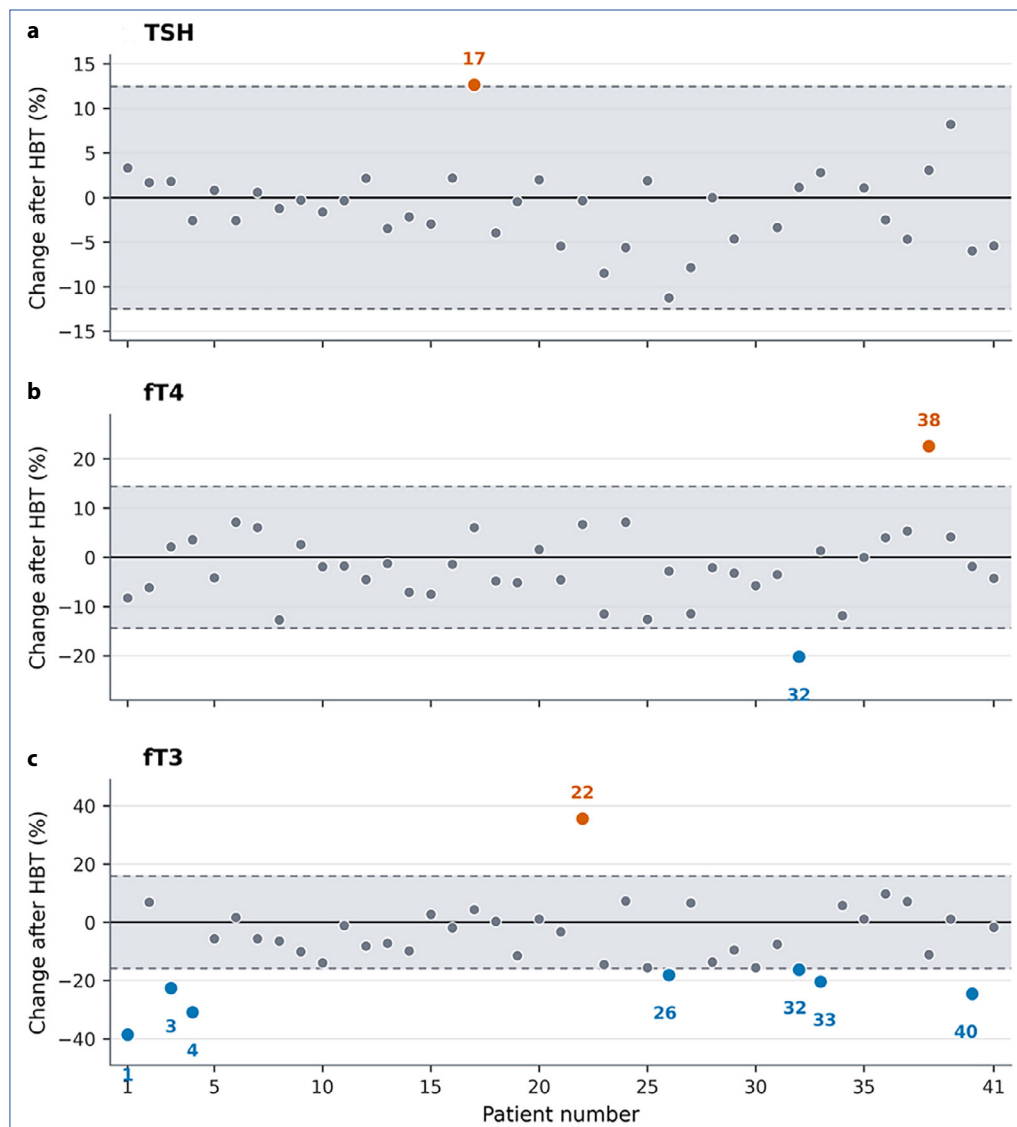


Figure 1. Percentage changes in thyroid function test results following heterophile-blocking treatment. Panels show the percentage changes in (a) thyroid-stimulating hormone (TSH), (b) free thyroxine (fT4), and (c) free triiodothyronine (fT3) for the 41 patients. Signed percentage change was calculated as $100 \times (\text{post-HBT} - \text{pre-HBT})/\text{pre-HBT}$. The shaded bands represent the analyte-specific analytical change limits (ACLs), calculated as $\pm 2.77 \times CV_a$. The CV_a values were 4.50% for TSH, 5.20% for fT4, and 5.72% for fT3, yielding ACLs of $\pm 12.47\%$, $\pm 14.40\%$, and $\pm 15.84\%$, respectively. Gray points represent changes within the corresponding ACL, whereas orange and blue points represent increases and decreases exceeding the ACL, respectively. Patient numbers are shown for results exceeding the corresponding ACL. Percentage changes in TSH were not calculated for patients 30 and 34 because both their pre-HBT and post-HBT results were reported as <0.015 mIU/L. Under a two-sided 95% ACL, approximately 5% of points (about two per panel) are expected to fall outside the band by chance, provided that the imprecision estimate reflects the paired measurements under the study conditions.

ACL: Analytical change limit; CV_a : Analytical coefficient of variation; fT3: Free triiodothyronine; fT4: Free thyroxine; HBT: Heterophile-blocking tube; TSH: Thyroid-stimulating hormone.

the study conditions. Observed exceedance was close to this nominal rate for TSH (1/39, 2.6%; $p=0.86$) and fT4 (2/41, 4.9%; $p=0.61$), but not for fT3 (8/41, 19.5%; $p=0.0008$) (Table 3). Exceedance differed significantly across the three analytes in the same 39 patients (Cochran's $Q=8.60$, $df=2$, $p=0.014$; exact $p=0.018$): 29 patients showed no exceedance, seven exceed-

ed the fT3 ACL only, one exceeded both the fT3 and fT4 ACLs, one exceeded the fT4 ACL only, and one exceeded the TSH ACL only. No systematic increase in TSH was observed: 23 of 39 quantifiable pairs decreased and 15 increased, and the single pair exceeding the TSH ACL remained within the reference interval. Percentage changes were not significantly correlated

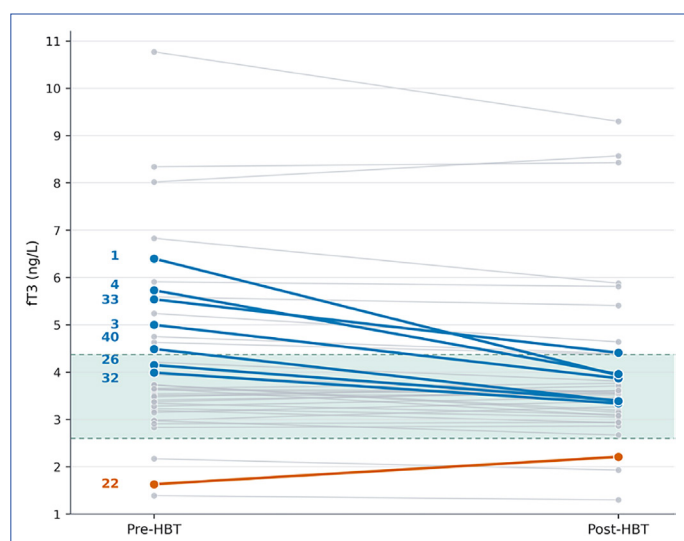


Figure 2. Individual changes in free triiodothyronine concentrations following heterophile-blocking treatment. Each line connects the pre-HBT and post-HBT ft3 measurements of an individual patient. The shaded band represents the laboratory reference interval (2.60–4.37 ng/L). Eight patients whose absolute percentage changes exceeded the ft3 ACL of 15.84% are highlighted and labeled. Blue lines indicate decreases and the orange line indicates an increase following HBT treatment; patients whose changes remained within the ACL are shown in gray. Among the eight highlighted patients, patients 1, 3, 4, and 40 showed post-HBT normalization with ACL exceedance, moving from above to within the reference interval. Patients 26 and 32 were already within the reference interval before HBT treatment, patient 33 remained above the upper limit, and patient 22 remained below the lower limit.

ACL: Analytical change limit; ft3: Free triiodothyronine; HBT: Heterophile-blocking tube.

across analytes within specimens (ft3 vs. ft4, $p=0.25$, $p=0.11$; ft3 vs. TSH, $p=-0.10$, $p=0.56$; ft4 vs. TSH, $p=0.17$, $p=0.29$). In the eight specimens with ft3 changes exceeding the ACL, concurrent TSH changes ranged from -11.2% to +3.3% (all within the TSH ACL), and ft4 changes ranged from -20.2% to +6.7%; only patient 32 also exceeded the ft4 ACL.

The ft3 change was inversely related to the pre-HBT ft3 concentration ($p=-0.45$, $p=0.003$). Among 14 specimens with pre-HBT ft3 above the upper reference limit, the median change was -12.5%, and 5 (35.7%) exceeded the ACL, whereas among 27 specimens with ft3 within or below the reference interval, the median change was -1.9% (paired $p=0.10$), and 3 (11.1%) exceeded the ACL. Among the 22 specimens within the reference interval that did not exceed the ACL, the median change was -0.4% ($p=0.22$). After exclusion of the eight ACL-exceeding specimens, a small residual decrease in ft3 persisted ($n=33$; median, -3.2%; IQR, -9.9% to +1.7%; $p=0.02$), similar in magnitude to the overall ft4 change. Across assumed CV_a values of 1.8–8% (ACL, 5.0–22.2%), ft3 exceedance ranged from 30/41 to 5/41 and was higher than TSH (9/39 to 0/39) and ft4 (19/41 to 1/41) exceedance at every assumption (Table 3); with the manufacturer's within-run CV_a range for ft3 (2.6–5.1%), 22 to 11 ft3 pairs would have exceeded the ACL.

Discussion

The principal finding was that changes exceeding the analyte-specific ACL after HBT treatment occurred predominantly in the ft3 assay: 8 of 41 pairs (19.5%) showed an absolute ft3 change greater than 15.84%, four (9.8%) both exceeded this limit and moved from above to within the reference interval, and ft3 was the only analyte with a significant paired difference after Holm adjustment. The ACL was exceeded in only 1 of 39 TSH pairs and 2 of 41 ft4 pairs, with no threshold-supported normalization, close to the approximately 5% expected for a two-sided 95% limit when no effect is present, provided that the imprecision estimate is appropriate, whereas ft3 exceedance was approximately four times higher. A response to blocking may thus be confined to one component of the thyroid profile even when all three analytes are measured in the same specimen on the same platform, consistent with the known dependence of antibody interference on assay format, reagents, and blocking strategy [1, 2, 5]. Current guidance recommends HBT pretreatment but provides no criterion for determining when a post-HBT change is analytically meaningful [5], and the criteria used in published blocking studies are heterogeneous, as discussed below. A limit derived in advance from the laboratory's own imprecision is transparent and transferable across laboratories, and the observed exceedance for analytes without a detectable HBT effect offers a simple, albeit assumption-dependent, plausibility check of the threshold.

The frequency of ft3 responses must be interpreted in the context of patient selection. Only 41 of 8,299 patients (0.49%) underwent HBT evaluation and were selected because their results were discordant with the clinical picture or with the remainder of the profile, in accordance with current recommendations [5]; the exceedance rate in this group therefore cannot be used to estimate interference in the general outpatient population, which would require a large unselected series [9, 25]. In a much larger study using Roche electrochemiluminescence assays, Imamura et al. [25] confirmed interference in 53 of 83,044 ft3 measurements (0.064%), compared with 11 of 124,609 TSH measurements (0.009%) and 6 of 116,012 ft4 measurements (0.005%). Although the populations, platforms, algorithms, and definitions differed, both studies identified ft3 as the most frequently affected analyte; the ft3 interferences characterized by Imamura et al. [25] were mainly attributed to idiosyncratic or anti-streptavidin antibodies, supporting a broader susceptibility of ft3 measurement rather than a specific predominance of classical heterophile antibodies.

The predominance of ft3 responses is unlikely to be explained by assay format alone: sandwich assays such as TSH are susceptible to antibody bridging, whereas competitive free-hormone assays may be affected by endogenous antibodies, altered binding proteins, tracer interactions, and other method-specific factors [1, 2]. In competitive assays, an endogenous antibody may bind the assay antibody, tracer or analog, or analyte itself and may therefore produce falsely low or falsely high results in the same assay; a nonuniform direc-

Table 2. Clinical and biochemical characteristics of patients with fT3 changes exceeding the analytical change limit after HBT treatment

Patient no.	Age, years	Recorded clinical context	Pre-HBT fT3, ng/L	Post-HBT fT3, ng/L	Signed change, %	fT3 reference-interval transition	Post-HBT TFT profile
1	31	Toxic multinodular goiter; methimazole discontinued	6.40	3.93	-38.6	Above to within ^a	TSH remained below the reference interval
3	31	Palpitations; propranolol use	5.00	3.87	-22.6	Above to within ^a	fT4 remained above the reference interval
4	35	Preoperative evaluation; previous HBT-responsive troponin result and bird exposure	5.73	3.96	-30.9	Above to within ^a	All three TFT results were within their respective reference intervals
22	74	Thyroid nodule; cardiovascular and metabolic comorbidities	1.63	2.21	+35.6	Below to below	fT4 remained above the reference interval
26	22	Previous fT4 elevation	4.15	3.40	-18.1	Within to within	All three TFT results remained within their respective reference intervals
32	23	Persistent TSH elevation; previous gastric cancer treatment	3.99	3.34	-16.3	Within to within	TSH remained above the reference interval; fT4 change (-20.2%) also exceeded the fT4 ACL
33	44	History of hypothyroidism; levothyroxine discontinued; microadenoma	5.54	4.41	-20.4	Above to above	TSH and fT4 remained above their respective reference intervals
40	33	Isolated fT3 elevation	4.49	3.39	-24.5	Above to within ^a	All three TFT results were within their respective reference intervals

^a: Post-HBT fT3 normalization with ACL exceedance was defined as a transition from above to within the fT3 reference interval (2.60–4.37 ng/L), accompanied by an absolute percentage change exceeding the fT3 ACL of 15.84%. Post-HBT normalization refers only to the fT3 result and does not imply clinical resolution. Concurrent TSH changes in these eight specimens ranged from -11.2% to +3.3% (all within the TSH ACL). ACL: Analytical change limit; fT3: Free triiodothyronine; fT4: Free thyroxine; HBT: Heterophile-blocking tube; TFT: Thyroid function test; TSH: Thyroid-stimulating hormone.

tion of change, as observed here (seven decreases and one increase), is thus expected rather than anomalous [1, 11]. Method dependence is well documented: Külz et al. [26] found that discordant fT4 and fT3 results varied among platforms, whereas total T3 remained within the reference interval in 70 of 71 samples; Cheng et al. [27] obtained TSH values ranging from 5.52 to <0.015 mIU/L for the same sample on four platforms, with markedly different responses to a blocking reagent; and Khoo et al. [28] found that fT4 assay susceptibility to familial dysalbuminaemic hyperthyroxinaemia, which was greatest for Beckman Access, depended on assay conditions rather than architecture. Because interference is method-specific and current guidance recommends that laboratories understand the behavior of the assays they use [5, 6], platform-specific data are required, and published HBT data for thyroid assays on the Dxl/Access platform are scarce. Whether the predominance of fT3 responses reflects a particular vulnerability of the Dxl fT3 assay, whose reagent formulation includes goat, bovine, and avian proteins and a streptavidin–biotin solid phase [22], remains uncertain and requires cross-platform evaluation.

Patient-level findings demonstrate why threshold exceedance and reference-interval transition should be evaluated separately. Of the eight patients with threshold-exceeding fT3 changes, four moved into the fT3 reference interval, but the complete profile normalized in only two; of the remaining

four, two were within the interval before and after treatment, one remained above it, and one increased but remained below it (Fig. 2, Table 2). Conversely, patient 7 reached the fT3 upper reference limit without exceeding the ACL, and two fT4 results normalized without exceeding the fT4 ACL, whereas two others moved outside the interval. Neither criterion alone confirms interference or establishes clinical significance.

No generally accepted decision rule exists for blocking experiments. Emerson et al. [19] classified a result as discrepant when $|Z|$, calculated from within-run imprecision, exceeded 4; Preissner et al. [20] used an empirical cutoff of 3 standard deviations from the mean percentage difference after HBT treatment. Our ACL of $2.77 \times CV_a$ is the two-sided 95% limit for the difference between paired measurements [23, 24]; because its basis and nominal cutoff differ from those criteria, response rates across studies should not be compared directly. Because the between-day CV_a was derived from quality-control data accumulated across runs and days, it may have yielded relatively conservative thresholds for same-run comparisons: with the manufacturer's within-run CV_a range for fT3, the number of ACL-exceeding fT3 results would have been 11–22 rather than 8, and the excess of fT3 over TSH and fT4 persisted at every CV_a assumption between 1.8% and 8%. Thus, although the absolute number of ACL-exceeding results depends on the imprecision estimate used, the excess

Table 3. Exploratory cross-analyte and sensitivity analyses of analytical change limit exceedance**A. Observed versus expected ACL exceedance in the same HBT-treated aliquots**

Analyte / subgroup	Exceeding ACL, n/N (%)	Nominal 5% of N	Exact binomial P	Signed change, %, median (IQR)
TSH	1/39 (2.6)	2.0	0.86	-0.45 (-3.71 to +1.74)
ft4	2/41 (4.9)	2.1	0.61	-2.11 (-5.77 to +2.59)
ft3	8/41 (19.5)	2.1	0.0008	-6.47 (-13.91 to +1.08)
ft3, pre-HBT ft3 > upper reference limit	5/14 (35.7)	0.7	—	-12.5
ft3, pre-HBT ft3 within or below reference interval	3/27 (11.1)	1.4	—	-1.9
ft3, within reference interval and not exceeding ACL	—	—	—	-0.4 (n = 22)
ft3, excluding ACL-exceeding specimens	—	—	—	-3.2 (-9.9 to +1.7)

B. Sensitivity of ACL classification to the assumed analytical coefficient of variation

Assumed CV _a	ACL (±%)	TSH exceeding, n/39	ft4 exceeding, n/41	ft3 exceeding, n/41
1.8%	5.0	9	19	30
2.6%	7.2	5	9	22
3.0%	8.3	3	7	19
4.0%	11.1	2	7	14
5.0%	13.9	0	2	12
5.1%	14.1	0	2	11
Laboratory between-day CV _a (4.50/5.20/5.72%)	12.5/14.4/15.8	1	2	8
7.0%	19.4	0	2	6
8.0%	22.2	0	1	5

Part A: the nominal proportion of pairs exceeding a two-sided 95% ACL in the absence of an effect is 5%, provided that the imprecision estimate reflects the paired measurements under the study conditions; P values are from the one-sided exact binomial test against 5% and are exploratory. Cochran's Q test for exceedance across the three analytes in the 39 patients with quantifiable pairs for all analytes: Q = 8.60, df=2, p=0.014 (exact permutation p=0.018); joint pattern: no exceedance in 29, ft3 only in 7, ft3 and ft4 in 1, ft4 only in 1, TSH only in 1. Spearman correlations of percentage changes across analytes: ft3 vs ft4, $\rho=0.25$ (p=0.11); ft3 vs TSH, $\rho=-0.10$ (p=0.56); ft4 vs TSH, $\rho=0.17$ (p=0.29). Part B: ACL=2.77 × assumed CV_a; the same assumed CV_a was applied to all three analytes except in the laboratory between-day row. Manufacturer's stated within-run CVs: TSH 2.5–3.1%, ft4 1.8–2.7%, ft3 2.6–5.1% at control concentrations within and above the reference intervals, and up to 6.6% at the lowest control level [22]; the 1.8–8% range encompasses all of these values. ACL: Analytical change limit; CV_a: Analytical coefficient of variation; df: Degrees of freedom; ft3: Free triiodothyronine; ft4: Free thyroxine; HBT: Heterophile-blocking tube; IQR: Interquartile range; TSH: Thyroid-stimulating hormone.

of ft3 exceedance over TSH and ft4 was present at every imprecision estimate examined. Our results also illustrate why a significant paired test should not be taken as evidence of interference: the cohort-level ft3 reduction was significant, yet its median magnitude (-6.5%) was well within the ACL, and only eight specimens exceeded the applied ACL. A significant group-level test can also arise from a small systematic shift present in every sample without any specimen being individually discrepant. Specimen-level criteria referenced to analytical imprecision [19] or to an empirically derived distribution of changes [20] are therefore required.

An HBT-associated change exceeding the ACL is compatible with endogenous antibody interference but does not confirm heterophile antibody interference or identify the interfering substance; blocking reagents do not neutralize every heterophile or human anti-animal antibody, and no single investigation can fully characterize interference [2, 5, 11]. Candidate mechanisms for HBT-responsive ft3 changes on this platform include heterophile and human anti-animal antibodies (the manufacturer notes that human anti-goat antibodies may interfere with the Free T3 assay [22]), thyroid hormone autoantibodies, and anti-idiotypic antibodies; the

present study cannot distinguish among them. Confirmatory approaches include remeasurement using a differently designed platform, total T3 measurement, and, for TSH, PEG precipitation or immunoglobulin depletion; equilibrium dialysis or ultrafiltration with mass spectrometry provides a more independent assessment of free hormones [29, 30], whereas serial dilution and PEG precipitation are less suitable for ft3 and ft4 because they disturb the bound-free equilibrium [2, 5, 11]. For suspected ft4/ft3 interference, the ETA guideline recommends an alternative assay, total hormone measurement, or referral to a specialist laboratory, while acknowledging that local availability may be limited [5]. The present study was designed to evaluate the HBT procedure as used in routine practice on the laboratory's single immunoassay platform; confirmation using a differently designed platform was outside the study protocol, and no alternative thyroid immunoassay or total T3 method was available at our institution. Although a liquid chromatography–tandem mass spectrometry system was available, direct free-hormone measurement requires prior separation of the free fraction by equilibrium dialysis or ultrafiltration, and no validated method of this kind had been established. The previous HBT-responsive troponin

result of patient 4, whose fT3 decreased by 30.9%, supports the clinical suspicion of an antibody-mediated mechanism in that patient but does not constitute independent analytical confirmation of the fT3 change (Table 2).

The questionnaire findings are descriptive: animal contact and recent infection were reported by 30.0% and 17.5% of patients with available information, respectively, but without a comparison group and with only eight threshold-exceeding fT3 results, the study could not assess whether any exposure was associated with interference. Heterophile reactivity occurs in individuals without recognized exposure [1, 7], so a negative exposure history does not exclude interference.

Although the study did not include a control group of samples without suspected interference subjected to the same HBT procedure, an internal reference was available within the study material. Preissner et al. [20] treated their entire study cohort and used the distribution of changes itself as the reference; a related approach was used here, with TSH and fT4 measured in the same treated aliquots serving as internal references. In our study, the cross-analyte comparisons addressed a nonspecific effect common to the entire aliquot: TSH and fT4 measured in the same treated aliquots exceeded their ACLs at rates close to the nominal rate, with a median TSH change of -0.45%, arguing against a volume or gross matrix effect of the tube. The clustering of fT3 changes in specimens with elevated pre-HBT fT3 is consistent with interference where it was clinically suspected, although this association is descriptive because percentage change is related to the baseline value. The small residual fT3 decrease after excluding the ACL-exceeding specimens (median, -3.2%) was similar to the overall fT4 change and remained within the ACL; it may represent subthreshold interference, a minor assay- or pretreatment-related effect, or both. HBT-related over-recovery has been described for Access assays with goat anti-mouse-coated particles [15, 16], and a marked HBT-associated increase in luteinizing hormone on the Dxl 800 was recently reported [21]. Of the three assays used here, only the HYPERsensitive hTSH assay has this design feature, and no systematic increase in TSH was observed; the fT3 and fT4 assays use streptavidin-biotin solid phases, so this particular mechanism is unlikely to explain the fT3 findings, although other reagent-related artifacts cannot be excluded. HBT applicability must therefore be verified assay by assay, and an HBT-treated control series on this platform is needed.

The strengths of the study include the paired same-specimen, same-run design, which excluded biological and between-run variation; the standardized blocking procedure; consistently applied analyte-specific limits evaluated alongside reference-interval transitions and individual trajectories; and the availability of TSH and fT4 results from the same treated aliquots. The main limitations are the small, single-center cohort; use of a single immunoassay platform; the absence of a separately treated control group, addressed in part by cross-analyte comparisons within the same treated aliquots; the absence of independent confirmation; reliance on between-day

rather than within-run CV_a; and incomplete questionnaire data. Where the data permitted, the influence of these limitations was examined rather than merely acknowledged through cross-analyte comparisons and sensitivity analysis. Without independent confirmation, the findings are more appropriately described as HBT-responsive changes compatible with endogenous antibody interference than as evidence of heterophile-antibody positivity, and they do not permit estimation of the prevalence of interference in the outpatient population or on the Dxl platform.

In conclusion, in this clinically selected cohort, an imprecision-based analytical change limit combined with reference-interval transition provided a transparent criterion for interpreting HBT results. HBT-associated changes exceeding the limit occurred predominantly in the fT3 assay, whereas comparable responses were uncommon for TSH and fT4 in the same treated aliquots. This pattern identifies fT3 on the Dxl platform as the analyte for which HBT results most often require further work-up and raises the possibility of method-specific susceptibility of fT3 measurement under the conditions studied, but it does not establish that every threshold-exceeding change was caused by a heterophile antibody or establish the mechanism or prevalence of interference. HBT treatment should remain one component of a broader interference investigation and should be interpreted in conjunction with the complete thyroid profile, clinical context, and, where necessary, independent confirmatory testing; close communication between clinicians and laboratory specialists remains essential when thyroid results are internally inconsistent or do not fit the clinical course.

Disclosures

Ethics Committee Approval: The study was approved by the Marmara University Faculty of Medicine Ethics Committee (no: 09.2025.25-0760, date: 05/09/2025).

Informed Consent: Informed consent was obtained from the participants.

Conflict of Interest Statement: None declared.

Funding: The authors declares that this study has received no financial support.

Use of AI for Writing Assistance: The authors declare that artificial intelligence-based tools were used solely for grammar checking and language editing during the preparation of this manuscript. All scientific content, data interpretation, and conclusions were produced by the authors in accordance with accepted scientific research methods and academic ethical principles.

Authorship Contributions: Concept – O.S., G.H.; Design – H.G., G.H.; Supervision – R.T., G.H.; Funding – H.G., O.S., G.H.; Materials – H.G., O.S., G.H.; Data collection and/or processing – R.T., T.C., A.Y.; Data analysis and/or interpretation – R.T., T.C., A.Y.; Literature search – R.T., T.C.; Writing – R.T., T.C.; Critical review – R.T., A.Y., T.C., H.G., O.S., G.H.

Peer-review: Externally peer-reviewed.

References

- Bolstad N, Warren DJ, Nustad K. Heterophilic antibody interference in immunometric assays. *Best Pract Res Clin Endocrinol Metab* 2013;27(5):647–61. [\[CrossRef\]](#)
- Favresse J, Burlacu MC, Maiter D, Gruson D. Interferences with thyroid function immunoassays: clinical implications and detection algorithm. *Endocr Rev* 2018;39(5):830–850. [\[CrossRef\]](#)
- Al-Bahadili H, Powers Carson J, Markov A, Jasim S. The complex web of interferences with thyroid function tests. *Endocr Pract* 2025;31(1):92–101. [\[CrossRef\]](#)
- Van Uytvanghe K, Ehrenkranz J, Halsall D, Hoff K, Loh TP, Spencer CA, et al. Thyroid stimulating hormone and thyroid hormones (triiodothyronine and thyroxine): An American Thyroid Association-commissioned review of current clinical and laboratory status. *Thyroid* 2023;33(9):1013–28. [\[CrossRef\]](#)
- Campi I, Feldt-Rasmussen U, Gruson D, Halsall D, Raverot V, van den Berg S, et al. 2026 ETA guideline on interference in immunoassay measurements used in assessment of thyroid function. *Eur Thyroid J* 2026;15(4):ETJ260011. [\[CrossRef\]](#)
- Ghazal K, Brabant S, Prie D, Piketty ML. Hormone immunoassay interference: a 2021 update. *Ann Lab Med* 2022;42(1):3–23. [\[CrossRef\]](#)
- Levinson SS, Miller JJ. Towards a better understanding of heterophile (and the like) antibody interference with modern immunoassays. *Clin Chim Acta* 2002;325(1-2):1–15. [\[CrossRef\]](#)
- Sun HG, Xu XP, Lu T, Yang Y, He LQ. A Pseudo Elevation of CEA Caused by Epidemic Hemorrhagic Fever. *Clin Lab* 2024;70(6). [\[CrossRef\]](#)
- Ismail AA, Walker PL, Barth JH, Lewandowski KC, Jones R, Burr WA. Wrong biochemistry results: two case reports and observational study in 5310 patients on potentially misleading thyroid-stimulating hormone and gonadotropin immunoassay results. *Clin Chem* 2002;48(11):2023–9. [\[CrossRef\]](#)
- Ismail AA. Interference from endogenous antibodies in automated immunoassays: what laboratorians need to know. *J Clin Pathol* 2009;62(8):673–8. [\[CrossRef\]](#)
- Sturgeon CM, Viljoen A. Analytical error and interference in immunoassay: minimizing risk. *Ann Clin Biochem* 2011;48(Pt 5):418–32. [\[CrossRef\]](#)
- Ward G, Simpson A, Boscatto L, Hickman PE. The investigation of interferences in immunoassay. *Clin Biochem* 2017;50(18):1306–11. [\[CrossRef\]](#)
- Wauthier L, Plebani M, Favresse J. Interferences in immunoassays: review and practical algorithm. *Clin Chem Lab Med* 2022;60(6):808–20. [\[CrossRef\]](#)
- Bolstad N, Warren DJ, Bjerner J, Kravdal G, Schwettmann L, Olsen KH, et al. Heterophilic antibody interference in commercial immunoassays; a screening study using paired native and pre-blocked sera. *Clin Chem Lab Med* 2011;49(12):2001–6. [\[CrossRef\]](#)
- Ellis MJ, Livesey JH. Techniques for identifying heterophile antibody interference are assay specific: study of seven analytes on two automated immunoassay analyzers. *Clin Chem* 2005;51(3):639–41. [\[CrossRef\]](#)
- Cantor D. Reason for limitations of heterophilic blocking tube use on certain Beckman Coulter Access assays. *Clin Chem* 2005;51(7):1311. [\[CrossRef\]](#)
- Çokluk E, Tuncer FB, Şekeroğlu MR, Gözükar C, Ağaç S, Gözükar Sİ. CA 19-9 ölçümünde heterofil antikör interferansı: bir olgu sunumu. *Sakarya Med J* 2019;9:558–64. [\[Article in Turkish\]](#). [\[CrossRef\]](#)
- Paczkowska K, Otlewska A, Loska O, Kolačkov K, Bolanowski M, Daroszewski J. Laboratory interference in the thyroid function test. *Endokrynol Pol* 2020;71(6):551–60. [\[CrossRef\]](#)
- Emerson JF, Ngo G, Emerson SS. Screening for interference in immunoassays. *Clin Chem* 2003;49(7):1163–9. [\[CrossRef\]](#)
- Preissner CM, O’Kane DJ, Singh RJ, Morris JC, Grebe SK. Phantoms in the assay tube: heterophile antibody interferences in serum thyroglobulin assays. *J Clin Endocrinol Metab* 2003;88(7):3069–74. [\[CrossRef\]](#)
- Ekinci A, Canpolat Erkan RE, Yildiz I, Misirlioglu NF, Uzun H. Use of heterophilic blocking tubes in suspected heterophile antibody interference among pubertal patients. *Medicina (Kaunas)* 2026;62(1):129. [\[CrossRef\]](#)
- Beckman Coulter, Inc. Access HYPERsensitive hTSH (ref. 33820; document A83808B), Access Free T3 (ref. A13422; document A38322D), and Access Free T4 (ref. 33880; document B01898B): instructions for use. Brea (CA): Beckman Coulter, Inc.; 2010–2013.
- Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. *Crit Rev Clin Lab Sci* 1989;27(5):409–37. [\[CrossRef\]](#)
- Fraser CG. Reference change values. *Clin Chem Lab Med* 2011;50(5):807–12. [\[CrossRef\]](#)
- Imamura C, Nishihara E, Fukata S, Ito M, Nishikawa M, Miyauchi A, et al. Interference in thyroid function tests using the electrochemiluminescence immunoassay. *Eur Thyroid J* 2026;15(2):ETJ250396. [\[CrossRef\]](#)
- Külz M, Fellner S, Rocktäschel J, Ceglarek U, Willenberg A, Kratzsch J. Dubiously increased FT4 and FT3 levels in clinically euthyroid patients: clinical finding or analytical pitfall? *Clin Chem Lab Med* 2022;60(6):877–85. [\[CrossRef\]](#)
- Cheng X, Guo X, Chai X, Hu Y, Lian X, Zhang G. Heterophilic antibody interference with TSH measurement on different immunoassay platforms. *Clin Chim Acta* 2021;512:63–5. [\[CrossRef\]](#)
- Khoo S, Lyons G, McGowan A, Gurnell M, Oddy S, Visser WE, et al. Familial dysalbuminaemic hyperthyroxinaemia interferes with current free thyroid hormone immunoassay methods. *Eur J Endocrinol* 2020;182(6):533–8. [\[CrossRef\]](#)
- Westbye AB, Aas FE, Kelp O, Dahl LK, Thorsby PM. Analysis of free, unbound thyroid hormones by liquid chromatography-tandem mass spectrometry: A mini-review of the medical rationale and analytical methods. *Anal Sci Adv* 2023;4(7-8):244–54. [\[CrossRef\]](#)
- Ma Z, Liu Z, Deng Y, Bai X, Zhou W, Zhang C. Free thyroid hormone: Methods and standardization. *Clin Chim Acta* 2025;565:119944. [\[CrossRef\]](#)

Appendix 1. Clinical history and exposure characteristics of patients evaluated for suspected immunoassay interference (n=41)

Characteristic	Value ^a
Age, years, median (IQR) [range]	41 (31–59) [22–74]
Female sex	27/41 (65.9%)
Male sex	14/41 (34.1%)
Known thyroid disease	20/38 (52.6%)
Thyroid medication use	17/38 (44.7%)
Non-thyroid medication use	22/38 (57.9%)
Upper respiratory tract infection or gastroenteritis within the previous 6 months	7/40 (17.5%)
Animal contact	12/40 (30.0%)
Allergic disease	5/41 (12.2%)
Influenza vaccination	5/41 (12.2%)
Previous blood or blood-product transfusion	2/41 (4.9%)

^a: Data are presented as median (interquartile range) [range] or n/N (%), where N represents the number of patients with available information. Clinical history and exposure variables were summarized descriptively. HBT: heterophile-blocking tube; IQR: Interquartile range.



Research Article

Internal validation and explainability of machine learning models for abnormal liver biochemistry

Osman Demir¹, Muzaffer Katar², Omer Faruk Nasip³

¹Department of Biostatistics, Tokat Gaziosmanpasa University Faculty of Medicine, Tokat, Türkiye

²Department of Biochemistry, Tokat Gaziosmanpasa University Faculty of Medicine, Tokat, Türkiye

³Department of Medical Education and Informatics, Tokat Gaziosmanpasa University Faculty of Medicine, Tokat, Türkiye

Abstract

Objectives: To develop, internally validate, and interpret machine learning models for classifying abnormal liver biochemistry using routine biochemical parameters.

Methods: This retrospective study included 335 individuals classified according to ICD-coded diagnoses retrieved from the hospital information system, comprising 175 individuals in the non-disorder group and 160 individuals in the abnormal liver biochemistry group. Aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, gamma-glutamyl transferase, total bilirubin, direct bilirubin, international normalized ratio, and albumin levels were used as predictors. Nine supervised learning models were evaluated using 10×5 repeated stratified cross-validation. The model with the numerically highest mean ROC-AUC was further examined using out-of-fold calibration analysis, exploratory decision curve analysis, and SHAP-based explanations.

Results: Aspartate aminotransferase showed the highest mean individual ROC-AUC (0.832), followed by alanine aminotransferase (0.796). XGBoost showed the numerically highest mean ROC-AUC (0.9640, 95% CI 0.9615–0.9664), with a mean accuracy of 0.9215 and F1-score of 0.9157. Mean sensitivity, specificity, positive predictive value, and negative predictive value were 0.9019, 0.9394, 0.9333, and 0.9161, respectively. Calibration analysis based on averaged repeated out-of-fold predictions yielded a Brier score of 0.056. SHAP analysis identified international normalized ratio, aspartate aminotransferase, and albumin as the most influential contributors to model predictions for the ICD-defined abnormal liver biochemistry outcome.

Conclusion: XGBoost showed high internally validated classification performance for the ICD-defined abnormal liver biochemistry outcome. These findings should be interpreted as internally validated classification performance rather than definitive clinical diagnostic accuracy. Although the outcome was based on ICD-coded diagnoses rather than biochemical threshold rules, external validation using independently adjudicated clinical outcomes remains necessary.

Keywords: Artificial intelligence, biomarkers, cross-validation studies, liver function tests, machine learning, ROC curve

How to cite this article: Demir O, Katar M, Nasip OF. Internal validation and explainability of machine learning models for abnormal liver biochemistry. Int J Med Biochem 2026;9(4):242–250.

The liver plays a central role in several essential physiological processes, including metabolism, detoxification, bile production, protein synthesis, and coagulation regulation [1, 2]. Abnormalities in liver-related biochemical parameters may indicate hepatocellular injury, cholestasis, impaired synthetic function, or mixed hepatic dysfunction [1]. As many liver disorders may remain asymptomatic or clinically nonspecific in

their early stages, routine biochemical tests are widely used as first-line indicators for detecting liver-related abnormalities and guiding further clinical evaluations [1].

Routine biochemical biomarkers, such as aspartate aminotransferase, alanine aminotransferase, alkaline phosphatase, gamma-glutamyl transferase, bilirubin, international normalized ratio, and albumin, provide complementary information

Address for correspondence: Osman Demir, PhD. Department of Biostatistics, Tokat Gaziosmanpasa University Faculty of Medicine, Tokat, Türkiye

Phone: +90 506 381 72 74 **E-mail:** osmand.demir@gop.edu.tr **ORCID:** 0000-0002-1322-2716

Submitted: July 13, 2026 **Revised:** August 14, 2026 **Accepted:** August 18, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



on different aspects of liver injury and function [1, 2]. Transaminases are commonly associated with hepatocellular injury, whereas alkaline phosphatase, gamma-glutamyl transferase, and bilirubin levels may reflect cholestatic or excretory abnormalities [1, 2]. Albumin and international normalized ratio are clinically relevant indicators of hepatic synthetic function and disease severity [2]. However, the interpretation of these biomarkers in combination may be challenging, as abnormal liver biochemistry may reflect complex and potentially nonlinear relationships among multiple biochemical parameters rather than an isolated abnormality in a single marker.

In recent years, machine learning methods have been increasingly applied to liver-related classification and computer-aided modeling tasks. Previous studies using clinical and biochemical datasets have compared algorithms such as logistic regression, k-nearest neighbors, decision tree, support vector machine, naïve Bayes, random forest, Extra Trees, and boosting-based models for classifying individuals with or without liver-related abnormalities [3, 4]. These studies commonly used biochemical variables such as bilirubin, alkaline phosphatase, aminotransferases, albumin, and albumin-to-globulin ratio, and reported model performance using accuracy, precision, recall, F1-score, specificity, and ROC-AUC [3, 4]. Systematic evidence also indicates that preprocessing, feature selection, outcome definition, and validation strategies are critical components of ML-based liver research [5].

Despite these advances, several methodological limitations persist. High apparent performance may occur when predictor variables overlap with the criteria used to define the outcome, creating a risk of incorporation bias or circular classification. Accuracy alone is insufficient for evaluating classification models, particularly when false-negative and false-positive classifications have different implications [6, 7]. A comprehensive evaluation should include discrimination, calibration, classification metrics, and internal validation [7–10]. Nevertheless, internal validation cannot establish transportability, and external validation is essential before broader clinical interpretation [8].

Another important issue is the interpretability of the results. Although complex machine learning algorithms may achieve high classification performance, their practical value is limited if the contribution of individual biomarkers to model predictions cannot be understood [11, 12]. Explainable artificial intelligence methods, particularly SHAP analysis, provide a framework for quantifying the relative contribution and direction of individual predictors [13].

Therefore, this study aimed to develop, internally validate, and interpret machine learning models for the classification of abnormal liver biochemistry using routine biochemical parameters. The study first evaluated the discriminative performance of individual biomarkers, then compared multiple supervised learning algorithms, and finally examined the calibration, exploratory decision-analytic net benefit, and SHAP-based explainability of the model with the numerically highest mean ROC-AUC. The analysis was intentionally positioned as an internally validated classification framework for an ICD-de-

fined abnormal liver biochemistry outcome, rather than an independently validated clinical diagnostic tool. We hypothesized that multivariable ensemble-based models would show higher internal discrimination than individual biomarkers and conventional single-model classifiers.

Materials and Methods

Study design and dataset

This retrospective classification study analyzed an existing dataset of 335 individuals, comprising 175 individuals in the non-disorder group and 160 in the ICD-defined abnormal liver biochemistry group. The analysis was conducted using anonymized laboratory data of the patients. No independent clinical diagnosis, imaging confirmation, disease stage, or etiological classification was available in the dataset.

The predictor variables were aspartate aminotransferase (AST; EC 2.6.1.1), alanine aminotransferase (ALT; EC 2.6.1.2), alkaline phosphatase (ALP; EC 3.1.3.1), gamma-glutamyl transferase (GGT; EC 2.3.2.2), total bilirubin, direct bilirubin, international normalized ratio (INR), and albumin levels. Enzyme activities were expressed as U/L, bilirubin concentrations as mg/dL, albumin as g/dL, and INR as a dimensionless ratio.

Group assignment was based on ICD-coded diagnoses retrieved from the hospital information system. Individuals were classified into the abnormal liver biochemistry group or the non-disorder group according to the diagnostic coding recorded in the hospital database. The outcome label was not generated by applying threshold-based rules to AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, or albumin values. Therefore, the binary outcome was not a deterministic function of the biochemical predictors evaluated in the present study.

Ethical considerations

The study was approved by the Tokat Gaziosmanpaşa University Non-Interventional Clinical Research Ethics Committee. (Date: 20.02.2026, Decision no: 26-MOBAEK-063). The study was conducted in accordance with the principles of the Declaration of Helsinki. The requirement for informed consent was waived because of the retrospective nature of the study. No other interventions were performed.

Definition of the abnormal liver biochemistry outcome

The binary outcome was based on ICD-coded diagnoses retrieved from the hospital information system and indicated membership in either the abnormal liver biochemistry group or the non-disorder group. Group assignment was therefore based on diagnostic coding rather than on threshold-based rules derived from the biochemical predictor variables. No deterministic combination of AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, or albumin was used to define the outcome. However, because laboratory findings may have contributed to the original clinical diagnostic process underlying ICD coding, residual incorporation bias cannot be completely excluded.

Data preprocessing

Before statistical analysis and model development, the dataset was examined for quality and consistency. Missing values, duplicate observations, outliers, variable formats, and class distributions were also assessed.

Missing values were identified only for INR (93/335, 27.8%), whereas the remaining biochemical predictors had complete data. Missing INR values were handled within the cross-validation procedure using median imputation. For each fold, the median was estimated exclusively from the training data and subsequently applied to the corresponding validation data to prevent information leakage. Continuous biochemical parameters, including AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, and albumin, were retained as numerical predictor variables. The outcome variable was encoded as a binary target variable.

The outcome classes were relatively balanced, with 175 and 160 individuals in the non-disorder and abnormal liver biochemistry groups, respectively. Therefore, no oversampling or synthetic data generation methods, such as SMOTE, were applied. To reduce the influence of scale differences among biochemical parameters, continuous predictors were standardized when required, particularly for scale-sensitive algorithms such as Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Multilayer Perceptron. Scaling parameters were estimated exclusively from the training data within each cross-validation fold and subsequently applied to the corresponding validation data to prevent information leakage.

Descriptive and comparative statistical analysis

Descriptive statistics were calculated for all biochemical parameters according to the study group. The distributional characteristics of continuous variables were assessed before group comparisons. Because the biochemical variables showed non-normal distributions, they were summarized as median and interquartile range (IQR). The non-disorder and abnormal liver biochemistry groups were compared in terms of AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, and albumin using the Mann–Whitney U test. A two-sided p-value of <0.05 was considered statistically significant.

Discriminative performance of individual biomarkers

The individual discriminative performance of each biochemical biomarker was evaluated using receiver operating characteristic (ROC) curve analysis. ROC analyses were performed separately for AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, and albumin to determine their ability to distinguish between the ICD-defined abnormal liver biochemistry and non-disorder groups.

Biomarker-specific performance was evaluated using the same 10×5 repeated stratified cross-validation framework. Within each fold, the optimal cutoff was determined from the training data using the Youden index (sensitivity + specificity

– 1) and subsequently applied to the corresponding validation data. ROC-AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the Youden index were calculated for the validation data and summarized across the repeated cross-validation procedure. For each biomarker, ROC-AUC was reported as the mean with a 95% confidence interval calculated from the 10 repetition-level mean ROC-AUC estimates using the t-distribution. For the individual INR analysis, only observations with available INR measurements were included, and no imputation was performed.

Conventional statistical modeling

Multivariable logistic regression analysis was performed using a conventional statistical reference model. The ICD-defined abnormal liver biochemistry outcome was used as the dependent variable. Because of substantial missingness in INR, the primary multivariable model included AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, and albumin as independent variables. AST, ALT, ALP, GGT, total bilirubin, and direct bilirubin were transformed using $\log(1+x)$ because of skewed distributions and subsequently standardized, whereas albumin was standardized without transformation. Odds ratios (ORs) were therefore reported per one-standard-deviation increase, together with 95% confidence intervals and p-values. Multicollinearity was assessed using variance inflation factors (VIFs). Sensitivity analyses were performed by excluding either total or direct bilirubin and by fitting a complete-case model including INR.

Model discrimination was assessed using ROC-AUC, overall probabilistic prediction error using the Brier score, and goodness of fit using the Hosmer–Lemeshow test. Nagelkerke R^2 was also calculated as a measure of model explanatory performance.

Machine learning model development and internal validation

Several supervised machine learning algorithms were developed to classify individuals according to the ICD-defined abnormal liver biochemistry outcome. The evaluated models included logistic regression, K-nearest neighbors, support vector machines, decision trees, random forests, gradient boosting, XGBoost, Naïve Bayes, and multilayer perceptron. No data-driven hyperparameter optimization was performed. Models were evaluated using fixed parameter settings within a uniform cross-validation framework to facilitate consistent model comparison and to limit optimization-related overfitting in the relatively small dataset. Detailed model-specific parameter settings are provided in Appendix 1.

Model performance was internally validated using 10×5 repeated stratified cross-validation (10 repetitions of five-fold cross-validation). Stratification was applied to preserve the distribution of the outcome classes in each fold. Performance metrics were averaged across the repeated cross-validation procedure. ROC-AUC was reported as the mean with standard deviation (SD) and 95% confidence interval (CI). For ROC-AUC, 95% confidence intervals were calculated from the 10 repe-

tition-level mean estimates using the t-distribution. Pairwise comparisons among the three models with the highest mean ROC-AUC values were performed using paired t-tests on the 10 matched repetition-level mean ROC-AUC estimates obtained from the repeated cross-validation procedure. Differences in mean ROC-AUC were reported with 95% confidence intervals and two-sided p-values.

Performance metrics

The primary performance metric was the ROC-AUC, which was used to assess the discriminatory ability of each model.

The secondary performance metrics included accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score. The abnormal liver biochemistry group was defined as the positive class for the calculation of these classification metrics. Confusion matrix-derived indices were calculated to provide an interpretable assessment of true-positive, true-negative, false-positive, and false-negative classifications.

Accuracy was defined as the proportion of correctly classified individuals among all individuals. Sensitivity represented the proportion of individuals in the abnormal liver biochemistry group who were correctly identified by the model, whereas specificity represented the proportion of individuals in the non-disorder group who were correctly classified. Positive predictive value (PPV) represented the proportion of predicted positive cases that were truly positive, whereas negative predictive value (NPV) represented the proportion of predicted negative cases that were truly negative. The F1-score was calculated as the harmonic mean of PPV and sensitivity.

Calibration, exploratory decision analysis, and explainability

For the model with the numerically highest mean ROC-AUC, additional analyses were performed to evaluate calibration, exploratory decision-analytic performance, and interpretability. Calibration was assessed using out-of-fold predicted probabilities generated during 10×5 repeated stratified cross-validation. For each individual, the out-of-fold predicted probabilities obtained across the 10 repetitions were averaged before calculation of the calibration curve and Brier score. The calibration curve was used for visual assessment, and the Brier score was calculated as a quantitative measure of overall probabilistic prediction error. Lower Brier scores indicate better overall probabilistic prediction accuracy.

Decision curve analysis was performed using the same averaged repeated out-of-fold predicted probabilities as an exploratory assessment of net benefit across a range of threshold probabilities. The model was compared with the default classify-all and classify-none strategies. Because the outcome was ICD-defined and no independent clinical endpoint was available, the decision curve analysis was considered exploratory and was not interpreted as evidence of established clinical utility.

To improve interpretability, SHAP analysis was performed for the model with the numerically highest mean ROC-AUC. SHAP values were calculated for validation observations within each cross-validation fold. For each individual, SHAP values obtained across the 10 repetitions were averaged to derive repeated out-of-fold feature contributions. A SHAP summary plot was generated to identify the most influential predictors and illustrate the direction and magnitude of individual biomarker contributions to the classification of the ICD-defined abnormal liver biochemistry outcome.

Statistical analysis

All analyses were performed using Python version 3.11. Data management was conducted using Pandas and NumPy, statistical analyses using SciPy and Statsmodels, machine learning using Scikit-learn and XGBoost, and explainability analysis using SHAP. Graphical outputs were generated using Matplotlib. Two-sided p-values <0.05 were considered statistically significant. Classification performance metrics were summarized as mean values across the repeated cross-validation procedure, while ROC-AUC was additionally reported with the standard deviation and 95% confidence interval.

The study was reported in accordance with the TRIPOD+AI reporting guidance [14], and the completed TRIPOD+AI checklist is provided as Appendix 2.

Results

Study population and dataset characteristics

The final dataset comprised 335 individuals, including 175 in the non-disorder group and 160 in the abnormal liver biochemistry group. These groups accounted for 52.2% and 47.8% of the dataset, respectively. The predictor variables included AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, and albumin.

The biochemical characteristics of the study groups are shown in Table 1.

All evaluated biochemical parameters differed significantly between the groups. The abnormal liver biochemistry group had higher median AST, ALT, ALP, GGT, total bilirubin, direct bilirubin, INR, and albumin values than the non-disorder group. INR comparisons were based on available cases without imputation.

Discriminative performance of individual biochemical biomarkers

The discriminative performance of individual biochemical biomarkers in distinguishing the abnormal liver biochemistry and non-disorder groups is presented in Table 2. Among the evaluated biomarkers, AST showed the highest mean ROC-AUC (0.832), followed by ALT (0.796). Across repeated cross-validation, the mean training-derived cutoff for AST was ≥ 23.32 , with a mean sensitivity of 77.4% and specificity of 78.4%. For ALT, the corresponding mean cutoff was ≥ 27.16 , with a mean sensitivity of 73.4% and specificity of 72.9%.

Table 1. Biochemical characteristics of the study groups

Variable	Non-disorder group (n=175)	Abnormal liver biochemistry (n=160)	p
AST	18.60 (15.00–22.45)	41.00 (24.00–66.78)	<0.001
ALT	16.50 (10.50–27.05)	51.80 (25.95–92.10)	<0.001
ALP	79.00 (64.00–106.00)	108.00 (81.00–176.00)	<0.001
GGT	21.00 (13.00–41.00)	42.00 (21.75–128.50)	<0.001
Total bilirubin	0.46 (0.30–0.69)	0.54 (0.36–0.81)	0.006
Direct bilirubin	0.16 (0.11–0.27)	0.21 (0.15–0.39)	<0.001
INR	1.04 (0.97–1.15)	1.08 (1.02–1.22)	0.009
Albumin	4.14 (3.24–4.58)	4.50 (4.32–4.70)	<0.001

Values are presented as median (interquartile range). Between-group comparisons were performed using the Mann-Whitney U test. Bilirubin conversion to SI units: mg/dL $\times 17.1 = \mu\text{mol/L}$; albumin conversion: g/dL $\times 10 = \text{g/L}$. INR descriptive statistics and between-group comparison were based on available cases without imputation. ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase; INR: International normalized ratio.

Table 2. Discriminative performance of individual biochemical biomarkers

Biomarker	ROC-AUC	95% CI	Mean cut-off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Youden index
AST	0.832	0.830–0.835	≥ 23.32	77.4	78.4	77.0	79.4	0.558
ALT	0.796	0.794–0.798	≥ 27.16	73.4	72.9	71.6	75.4	0.464
ALP	0.685	0.683–0.687	≥ 85.76	70.4	55.8	59.5	68.7	0.262
GGT	0.677	0.675–0.680	≥ 34.16	63.1	64.0	61.9	66.0	0.271
Total bilirubin	0.587	0.584–0.590	≥ 0.55	53.4	62.8	56.5	59.8	0.162
Direct bilirubin	0.638	0.635–0.641	≥ 0.19	65.4	58.8	59.1	65.5	0.242
INR	0.610	0.605–0.615	≥ 1.02	73.1	43.1	32.9	81.6	0.162
Albumin	0.658	0.656–0.660	≥ 4.45	69.4	64.9	64.6	70.1	0.343

Values represent mean performance across 10 $\times$ 5 repeated stratified cross-validation. Biomarker-specific cutoff values were determined in each training fold using the Youden index and evaluated in the corresponding validation fold. The positive outcome was the abnormal liver biochemistry group. INR analysis was based on available cases without imputation. ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; CI: Confidence interval; GGT: Gamma-glutamyl transferase; INR: International normalized ratio; NPV: Negative predictive value; PPV: Positive predictive value; ROC-AUC: Receiver operating characteristic area under the curve.

Compared with AST and ALT, the remaining biomarkers showed lower individual discriminative performance, with mean ROC-AUC values ranging from 0.587 to 0.685. Overall, individual biomarkers showed lower discrimination than the best-performing multivariable machine learning models.

Multivariable logistic regression analysis

Multivariable logistic regression analysis was performed to evaluate the associations between routine biochemical biomarkers and the ICD-defined abnormal liver biochemistry outcome. The results are presented in Table 3. AST, ALP, and albumin levels were significantly associated with group membership. In contrast, ALT, GGT, total bilirubin, and direct bilirubin were not statistically significant in the multivariable model.

Although VIF analysis indicated notable multicollinearity between total and direct bilirubin, sensitivity analyses excluding either bilirubin variable did not materially alter the associations of AST, ALP, and albumin with the outcome.

In a complete-case sensitivity analysis including INR (n=242), INR was not independently associated with the outcome (OR per 1-SD increase=1.046, 95% CI 0.699–1.565; p=0.828), while the principal associations for AST, ALP, and albumin were retained.

Table 3. Multivariable logistic regression analysis for the abnormal liver biochemistry outcome

Variable	OR per 1-SD increase	95% CI	p	VIF
AST	6.913	3.690–12.949	<0.001	2.42
ALT	1.519	0.919–2.513	0.103	2.94
ALP	1.796	1.166–2.767	0.008	1.67
GGT	0.999	0.630–1.585	0.998	2.03
Total bilirubin	1.027	0.492–2.143	0.944	7.63
Direct bilirubin	1.094	0.465–2.573	0.837	8.73
Albumin	2.600	1.765–3.829	<0.001	1.32

AST, ALT, ALP, GGT, total bilirubin, and direct bilirubin were transformed using log(1+x) because of skewed distributions and subsequently standardized. Albumin was standardized without transformation. Odds ratios therefore represent the change in odds associated with a one-standard-deviation increase in each predictor. INR was not included in the primary multivariable model because of substantial missingness. ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; CI: Confidence interval; GGT: Gamma-glutamyl transferase; OR: Odds ratio; VIF: Variance inflation factor.

The conventional logistic regression model showed an apparent ROC-AUC of 0.8854 and a Brier score of 0.1321 when fitted to the full analytical dataset. The Hosmer–Lemeshow test did not indicate significant lack of fit ($\chi^2=11.509$, df=8; p=0.175), and the Nagelkerke R^2 was 0.550.

Table 4. Cross-validated performance of machine learning models

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score	ROC-AUC, mean±SD (95% CI)
Logistic regression	0.7684	0.6937	0.8366	0.8000	0.7514	0.7401	0.8445±0.0066 (0.8397–0.8492)
KNN	0.7919	0.7669	0.8149	0.7934	0.7953	0.7778	0.8573±0.0118 (0.8488–0.8657)
SVM	0.7719	0.6812	0.8549	0.8128	0.7490	0.7383	0.8602±0.0057 (0.8561–0.8643)
Decision tree	0.8382	0.8400	0.8366	0.8297	0.8546	0.8322	0.8383±0.0149 (0.8276–0.8489)
Random forest	0.8979	0.8456	0.9457	0.9354	0.8727	0.8868	0.9539±0.0061 (0.9495–0.9583)
Gradient boosting	0.9185	0.8956	0.9394	0.9333	0.9102	0.9126	0.9615±0.0032 (0.9592–0.9639)
XGBoost	0.9215	0.9019	0.9394	0.9333	0.9161	0.9157	0.9640±0.0034 (0.9615–0.9664)
Naive bayes	0.6743	0.3956	0.9291	0.8407	0.6281	0.5342	0.8303±0.0040 (0.8274–0.8331)
Multilayer perceptron	0.8170	0.8006	0.8320	0.8176	0.8243	0.8061	0.8928±0.0076 (0.8873–0.8982)

KNN: K-nearest neighbors; MLP: Multilayer perceptron; NPV: Negative predictive value; PPV: Positive predictive value; ROC-AUC: Receiver operating characteristic area under the curve; SD: Standard deviation; SVM: Support vector machine; XGBoost: Extreme gradient boosting. Performance estimates were obtained using 10×5 repeated stratified cross-validation. Accuracy, sensitivity, specificity, PPV, NPV, and F1-score are presented as mean values, whereas ROC-AUC is presented as mean±SD (95% CI). Sensitivity, specificity, PPV, NPV, and F1-score were calculated with the abnormal liver biochemistry group as the positive class. Missing INR values were imputed within each training fold using the median estimated from that training fold and subsequently applied to the corresponding validation fold.

Cross-validated machine learning model performance

The cross-validated performance of the evaluated machine learning models is presented in Table 4. Among the evaluated models, XGBoost showed the numerically highest mean ROC-AUC (0.9640, 95% CI 0.9615–0.9664), with a mean accuracy of 0.9215 and F1-score of 0.9157. Gradient Boosting and Random Forest also showed high mean ROC-AUC values of 0.9615 and 0.9539, respectively. These estimates reflect internal classification performance for the abnormal liver biochemistry outcome.

Pairwise comparisons showed no statistically significant difference in mean ROC-AUC between XGBoost and Gradient Boosting (Δ ROC-AUC=0.0024, 95% CI –0.0002 to 0.0051; $p=0.069$). XGBoost showed a significantly higher mean ROC-AUC than Random Forest (Δ ROC-AUC=0.0101, 95% CI 0.0066–0.0135; $p<0.001$), while Gradient Boosting also showed a significantly higher mean ROC-AUC than Random Forest (Δ ROC-AUC=0.0077, 95% CI 0.0036–0.0117; $p=0.002$).

Tree-based ensemble and boosting algorithms generally showed higher observed classification performance than conventional and single-model classifiers. However, these findings represent internal discrimination of the ICD-defined abnormal liver biochemistry outcome and do not establish independent clinical diagnostic validity.

ROC-AUC performance of the machine learning models

Figure 1 shows the mean receiver operating characteristic (ROC) curves of the evaluated machine learning models across 10×5 repeated stratified cross-validation. Consistent with the repeated cross-validation results presented in Table 4, XGBoost showed the numerically highest mean ROC-AUC (0.9640), followed by Gradient Boosting (0.9615) and Random Forest (0.9539). Among the remaining models, Multilayer Perceptron showed the highest mean ROC-AUC (0.8928), followed by SVM (0.8602) and KNN (0.8573).

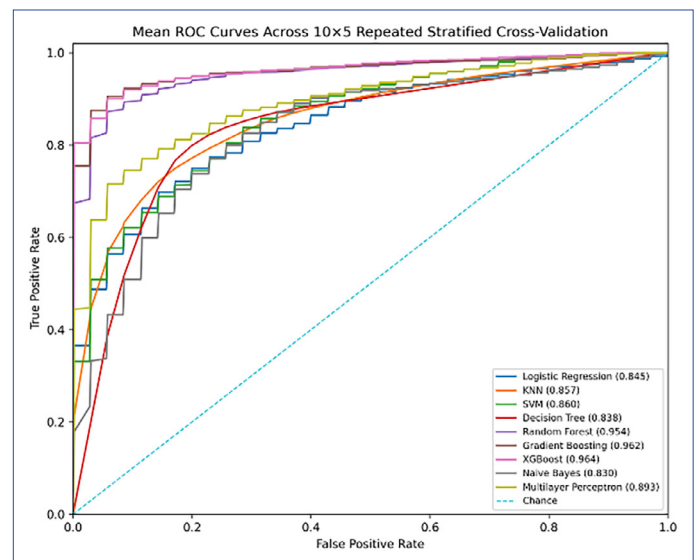


Figure 1. Mean receiver operating characteristic curves of the evaluated machine learning models across 10×5 repeated stratified cross-validation. XGBoost showed the numerically highest mean ROC-AUC, followed by Gradient Boosting and Random Forest.

GBM: Gradient boosting machine; KNN: K-nearest neighbors; LR: Logistic regression; MLP: Multilayer perceptron; NB: Naïve Bayes; RF, random forest; ROC-AUC: Receiver operating characteristic area under the curve; SVM: Support vector machine; XGBoost: Extreme gradient boosting.

Overall, the ensemble-based tree models showed higher observed internal discrimination than the conventional and single-classifier models. Therefore, XGBoost was selected for additional calibration, exploratory decision curve, and explainability analyses.

Cross-validated classification performance of the selected XGBoost model

Based on its numerically highest mean ROC-AUC among the evaluated models, XGBoost was selected for additional calibration, exploratory decision curve, and explainability analyses. Across 10×5 repeated stratified cross-validation, XGBoost

achieved a mean accuracy of 0.9215, sensitivity of 0.9019, specificity of 0.9394, positive predictive value of 0.9333, negative predictive value of 0.9161, F1-score of 0.9157, and ROC-AUC of 0.9640 (95% CI 0.9615–0.9664).

For descriptive visualization of classification errors, an aggregate out-of-fold confusion matrix was constructed by averaging each individual's predicted probabilities across the 10 cross-validation repetitions and applying a classification threshold of 0.50. The resulting matrix included 166 true-negative, 9 false-positive, 13 false-negative, and 147 true-positive classifications (Appendix 3). Because this aggregate matrix was based on patient-level averaged out-of-fold predictions, its derived classification metrics are not directly equivalent to the mean fold-level performance estimates reported in Table 4.

These estimates represent internally validated classification performance for the ICD-defined abnormal liver biochemistry outcome and should not be interpreted as independent clinical diagnostic accuracy.

Calibration, exploratory decision analysis, and explainability

Calibration analysis based on averaged repeated out-of-fold predictions yielded a Brier score of 0.056, indicating relatively low overall probabilistic prediction error. Exploratory decision curve analysis based on the same averaged repeated out-of-fold predictions showed that the model provided a higher net benefit than the classify-all and classify-none strategies across a broad range of threshold probabilities. Because the outcome was ICD-defined and no independent clinical endpoint was available, these findings should be interpreted as exploratory decision-analytic results rather than evidence of established clinical utility (Appendix 4, 5).

SHAP analysis was used to evaluate the contribution and direction of biochemical variables in the XGBoost model. The SHAP summary plot is shown in Figure 2. INR had the highest mean absolute SHAP value, followed closely by AST and then albumin, indicating that these variables contributed most strongly to model predictions. These findings reflect the relative contribution of the variables to the classification of the ICD-defined abnormal liver biochemistry outcome but do not establish causal or pathophysiological importance.

Discussion

This study developed and internally validated machine learning models for classifying abnormal liver biochemistry using routine biochemical parameters. Internal validation was performed using 10×5 repeated stratified cross-validation. Individual biomarkers showed modest discriminatory performance, whereas multivariable machine learning models showed higher internal classification performance. Among the evaluated models, XGBoost showed the numerically highest mean ROC-AUC and was selected for calibration, exploratory decision analysis, and SHAP-based interpretation. These findings reflect the model's ability to classify an ICD-defined

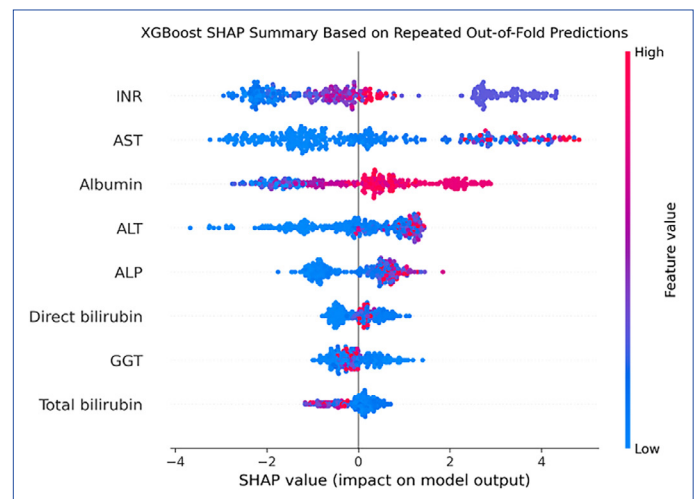


Figure 2. Repeated out-of-fold SHAP summary plot of the XGBoost model. SHAP values were calculated for validation observations within each cross-validation fold and averaged across the 10 repetitions for each individual. Each point represents an individual observation. Features are ranked according to their mean absolute SHAP values. Higher feature values are shown in red and lower values in blue. Positive SHAP values increase, whereas negative SHAP values decrease, the predicted probability of the ICD-defined abnormal liver biochemistry outcome.

ALP: Alkaline phosphatase; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; GGT: Gamma-glutamyl transferase; INR: International normalized ratio; SHAP: SHapley Additive exPlanations.

outcome using routine biochemical predictors and should not be interpreted as evidence of independently validated clinical diagnostic performance [14].

The between-group differences in AST, ALT, ALP, GGT, bilirubin, INR, and albumin levels are consistent with the established clinical roles of these parameters in characterizing liver-related laboratory abnormalities [1, 2]. AST and ALT showed the highest individual ROC-AUC values, indicating greater individual discrimination between the ICD-defined abnormal liver biochemistry and non-disorder groups than the other evaluated biomarkers. However, although the group label was derived from ICD-coded diagnoses rather than directly from biochemical thresholds, laboratory abnormalities may have contributed to the original diagnostic coding process; therefore, some degree of incorporation bias cannot be entirely excluded.

Although AST and ALT showed the strongest individual discrimination, their ROC-AUC values were lower than those of the multivariable machine learning models. This finding suggests that combining information across multiple biochemical parameters may improve discrimination compared with individual biomarkers alone. Similar ensemble-based approaches have demonstrated strong performance in liver-related datasets [15, 16]. Nevertheless, the apparent advantage of multivariable modeling in the present study may partly reflect associations between the biochemical predictors and the ICD-defined outcome, particularly because laboratory findings may have contributed to the original diagnostic coding process.

Among the evaluated algorithms, XGBoost showed the numerically highest mean ROC-AUC and balanced classification metrics. However, the difference in mean ROC-AUC between XGBoost and Gradient Boosting was not statistically significant, indicating that the observed numerical advantage of XGBoost over Gradient Boosting should not be interpreted as evidence of clear statistical superiority. Its performance may be related to its ability to model nonlinear relationships and interactions through regularized gradient boosting [17]. Nevertheless, increasing algorithmic complexity cannot overcome bias introduced by the definition of the outcome variable. Accordingly, the high performance observed in this study should be interpreted as internally validated classification performance for an ICD-defined outcome rather than as evidence that XGBoost can independently diagnose liver disease in other clinical populations.

Calibration and decision-analytic assessments provided complementary information beyond discrimination. Calibration based on averaged repeated out-of-fold predictions yielded a Brier score of 0.056, although the calibration curve deviated from the ideal line in the intermediate probability ranges. Exploratory decision curve analysis, based on the same out-of-fold predictions, showed a higher net benefit than the classify-all and classify-none strategies across a broad range of threshold probabilities. Because the endpoint was based on ICD-coded diagnoses and external validation was not performed, these results should not be interpreted as proof of clinical usefulness; rather, they describe potential decision-analytic behavior within the current dataset [7, 9, 10].

SHAP analysis improved the model transparency by showing how each biochemical parameter contributed to individual predictions. INR had the highest mean absolute SHAP value, followed closely by AST and then albumin, indicating that these variables had the largest overall contributions to the XGBoost predictions. These results help explain the internal logic of the model; however, SHAP importance should not be interpreted as causality or independent clinical importance, particularly because laboratory findings may have contributed to the clinical diagnostic process underlying the ICD-coded outcome [11, 13].

This study had several strengths. We compared individual biomarker discrimination with multivariable machine learning classification, used 10×5 repeated stratified cross-validation for internal validation, and evaluated discrimination, calibration, exploratory decision-analytic net benefit, and explainability. The study also provides a transparent account of the limitations imposed by an ICD-defined abnormal liver biochemistry outcome and positions the analysis as an exploratory, internally validated classification framework.

This study has some important limitations. An important limitation is the potential for residual incorporation bias. Although the outcome label was based on ICD-coded diagnoses retrieved from the hospital information system and was not generated directly from the biochemical predictors, laboratory findings may have contributed to the original clinical

diagnostic process and subsequent ICD coding. Moreover, the analytical dataset did not contain independently adjudicated clinical outcomes or detailed information on the diagnostic process underlying each ICD-coded classification. Consequently, the observed model performance should be interpreted as internally validated classification performance for an ICD-defined outcome and not as validation against an independently adjudicated clinical endpoint.

The dataset was relatively small, retrospective, and derived from a single center, and external validation was not performed. The relatively small sample size in relation to the number and complexity of the evaluated algorithms may also increase the risk of model-selection optimism, despite the use of repeated cross-validation. Although 10×5 repeated stratified cross-validation was used to improve the robustness of internal performance estimates, this procedure does not substitute for validation in an independent external population. No data-driven hyperparameter optimization was performed; therefore, the reported results reflect the performance of the prespecified model configurations rather than optimized versions of the evaluated algorithms. Detailed clinical information on etiology, disease stage, imaging findings, comorbidities, medication use, and longitudinal outcomes was unavailable. Albumin-related findings should be interpreted cautiously because albumin levels were unexpectedly higher in the abnormal liver biochemistry group, which may reflect characteristics of the study population or outcome definition rather than a typical disease-related pattern. The non-disorder group should therefore not be interpreted as a prospectively recruited healthy reference population, and unmeasured differences in clinical setting or patient selection may have contributed to the observed albumin distribution. Finally, the calibration curve showed some deviation from ideal calibration, particularly in the intermediate probability range. Therefore, external validation using independently and clinically adjudicated outcomes is essential before considering clinical application.

Conclusion

In conclusion, routine biochemical parameters supported the internal classification of the ICD-defined abnormal liver biochemistry outcome. Among the evaluated models, XGBoost showed the numerically highest mean ROC-AUC and was selected for additional calibration, exploratory decision analysis, and explainability analyses. Repeated out-of-fold SHAP analysis identified INR, followed closely by AST and albumin, as the most influential contributors to XGBoost predictions.

This study should be interpreted as an exploratory, internally validated classification analysis of an ICD-defined outcome rather than as a clinical diagnostic model. The possibility of incorporation bias, absence of independent outcome adjudication, limited sample size, and lack of external validation restrict clinical interpretation. Future studies should evaluate the model in larger multicenter cohorts using independently defined clinical outcomes before considering any diagnostic or decision-support applications.

Disclosures

Ethics Committee Approval: The study was approved by the Tokat Gaziosmanpaşa University Non-Interventional Clinical Research Ethics Committee (no: 26-MOBAEK-063, date: 20/02/2026).

Informed Consent: The requirement for informed consent was waived by the ethics committee because of the retrospective study design and the use of existing anonymized data.

Conflict of Interest Statement: None declared.

Funding: The author declares that this study has received no financial support.

Use of AI for Writing Assistance: OpenAI ChatGPT was used for language editing, wording refinement, and formatting assistance. The authors critically reviewed and revised all AI-assisted content and take full responsibility for the accuracy, integrity, and final content of the manuscript.

Authorship Contributions: Concept – O.D., M.K.; Design – O.D., M.K., O.F.N.; Supervision – O.D.; Funding – M.K.; Materials – M.K.; Data collection and/or processing – M.K.; Data analysis and/or interpretation – O.D., M.K., O.F.N.; Literature search – O.D., O.F.N.; Writing – O.D.; Critical review – M.K., O.F.N.

Peer-review: Externally peer-reviewed.

References

1. Kwo PY, Cohen SM, Lim JK. ACG clinical guideline: evaluation of abnormal liver chemistries. *Am J Gastroenterol* 2017;112:18–35. [\[CrossRef\]](#)
2. Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, et al. Guidelines on the management of abnormal liver blood tests. *Gut* 2018;67:6–19. [\[CrossRef\]](#)
3. Mostafa F, Hasan E, Williamson M, Khan H. Statistical machine learning approaches to liver disease prediction. *Livers* 2021;1(4):294–312. [\[CrossRef\]](#)
4. Rabbi MF, Mahedy Hasan SM, Champa AI, Asifzaman M, Hasan MK. Prediction of liver disorders using machine learning algorithms: a comparative study. *Proc Int Conf Artificial Intelligence and Computational Technologies* 2020:111–6. [\[CrossRef\]](#)
5. Khan RA, Luo Y, Wu FX. Machine learning based liver disease diagnosis: a systematic review. *Neurocomputing* 2022;468:492–509. [\[CrossRef\]](#)
6. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861–74. [\[CrossRef\]](#)
7. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21(1):128–38. [\[CrossRef\]](#)
8. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594. [\[CrossRef\]](#)
9. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group ‘Evaluating diagnostic tests and prediction models’ of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17(1):230. [\[CrossRef\]](#)
10. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–74. [\[CrossRef\]](#)
11. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Benetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence: concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 2020;58:82–115. [\[CrossRef\]](#)
12. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of any classifier. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2016:1135–44. [\[CrossRef\]](#)
13. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst* 2017;30:4765–74.
14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. [\[CrossRef\]](#)
15. Dalal S, Onyema EM, Malik A. Hybrid XGBoost model with hyperparameter tuning for prediction of liver disease with better accuracy. *World J Gastroenterol* 2022;28(46):6551–63. [\[CrossRef\]](#)
16. Lu CH, Wang W, Li YJ, Chang IW, Chen CL, Su CW, et al. Machine learning models for predicting significant liver fibrosis in patients with severe obesity and nonalcoholic fatty liver disease. *Obes Surg* 2024;34(12):4393–404. [\[CrossRef\]](#)
17. Chen T, Guestrin X. XGBoost: a scalable tree boosting system. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min* 2016:785–94. [\[CrossRef\]](#)

Appendix 1. Model-specific parameter settings

Model	Preprocessing within each cross-validation fold	Model-specific settings
Logistic regression	Median imputation; StandardScaler	C=1.0; solver=lbfgs; max_iter=1000; class_weight=None; random_state=42; penalty was not explicitly set in the estimator call (scikit-learn default behavior)
K-nearest neighbors (KNN)	Median imputation; StandardScaler	n_neighbors=5; weights=uniform; algorithm=auto; metric=Minkowski; p=2
Support vector machine (SVM)	Median imputation; StandardScaler	C=1.0; kernel=RBF; gamma=scale; probability=True; class_weight=None; random_state=42
Decision tree	Median imputation; no scaling	criterion=Gini; splitter=best; max_depth=None; min_samples_split=2; min_samples_leaf=1; class_weight=None; random_state=42
Random forest	Median imputation; no scaling	n_estimators=100; criterion=Gini; max_depth=None; min_samples_split=2; min_samples_leaf=1; max_features=sqrt; bootstrap=True; class_weight=None; random_state=42
Gradient boosting	Median imputation; no scaling	n_estimators=100; learning_rate=0.1; max_depth=3; subsample=1.0; max_features=None; random_state=42
XGBoost	Median imputation; no scaling	booster=gbtrees; n_estimators=100; learning_rate=0.3; max_depth=6; min_child_weight=1; subsample=1.0; colsample_bytree=1.0; gamma=0; reg_alpha=0; reg_lambda=1; objective=binary:logistic; eval_metric=logloss; random_state=42
Naïve bayes	Median imputation; no scaling	GaussianNB; var_smoothing=1×10 ⁻⁹
Multilayer perceptron (MLP)	Median imputation; StandardScaler	hidden_layer_sizes=(100,); activation=ReLU; solver=Adam; alpha=0.0001; learning_rate=constant; learning_rate_init=0.001; max_iter=5000; early_stopping=False; random_state=42

No data-driven hyperparameter tuning was performed. Missing INR values were imputed using the median estimated exclusively from each training fold and applied to the corresponding validation fold. Standardization, where indicated, was likewise fitted only on the training fold. Internal validation used RepeatedStratifiedKFold with 5 folds, 10 repeats, and random_state=42. The settings shown are the exact estimator arguments used in the verified rerun that reproduced the reported Table 4 mean performance values. Software versions in the verified rerun were scikit-learn 1.8.0 and XGBoost 3.1.3.

Appendix 2. TRIPOD+AI reporting checklist [14]

Item	Topic	How addressed in this study	Reported on page/section
1	Title	Development and internal validation, target context, and abnormal liver biochemistry outcome are identified.	Title
2	Abstract	Objectives, retrospective data source, predictors, internal validation, performance, calibration, and interpretation are summarized.	Abstract
3a	Background	Clinical context, rationale, and previous machine-learning studies in liver-related classification are described.	Introduction
3b	Target population/intended purpose	The study is positioned as an internally validated laboratory classification framework rather than an independently validated clinical diagnostic tool.	Introduction
3c	Health inequalities	Sociodemographic variables required for assessment of health inequalities were not available in the dataset.	Not applicable
4	Objectives	Development, internal validation, model comparison, calibration, exploratory decision analysis, and explainability objectives are stated.	Introduction
5a	Data source	An existing retrospective anonymized laboratory dataset and ICD-coded diagnoses from the hospital information system were used.	Methods: Study design and dataset
5b	Data dates	The retrospective dataset was obtained from the institutional hospital information system; specific accrual dates were not available in the analytical dataset.	Methods: Study design and dataset
6a	Setting / centres	The study used retrospective data derived from a single institutional setting.	Methods; Discussion: Limitations
6b	Eligibility criteria	The final analytical dataset comprised 335 individuals classified into non-disorder and abnormal liver biochemistry groups according to recorded ICD-coded diagnoses.	Methods: Study design and dataset
6c	Treatments	Treatment and medication information was unavailable and was not used for model development.	Discussion: Limitations

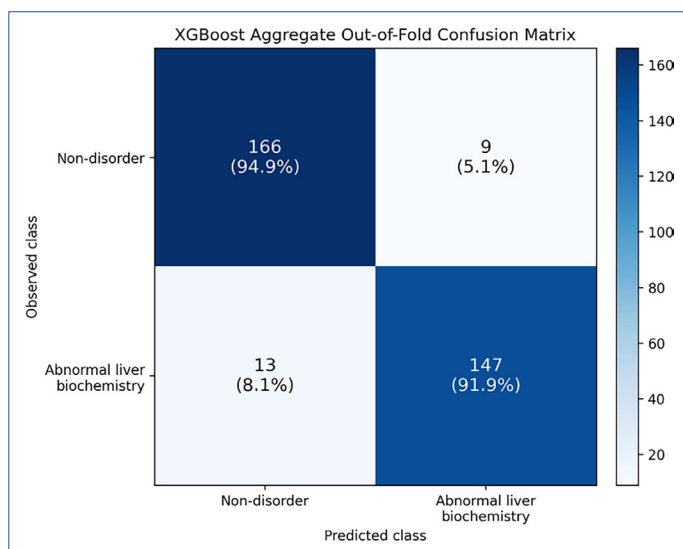
Appendix 2. Cont.

Item	Topic	How addressed in this study	Reported on page/section
7	Data preparation	Data-quality checks, missing-data handling, fold-wise imputation, scaling, class distribution, and procedures to prevent information leakage are described.	Methods: Data preprocessing
8a	Outcome	The binary outcome was based on ICD-coded diagnoses and was not generated using deterministic biochemical threshold rules.	Methods: Definition of the abnormal liver biochemistry outcome
8b	Outcome assessor subjectivity	Outcome labels were based on pre-existing ICD coding; independent clinical adjudication was unavailable.	Methods; Discussion
8c	Outcome blinding	No study-level outcome adjudication was performed because outcome labels pre-existed in the hospital information system.	Methods
9a	Predictor selection	Eight routinely available biochemical parameters were evaluated as predictors.	Introduction; Methods
9b	Predictor definition / measurement	Predictors, measurement units, and handling procedures are described.	Methods: Study design and dataset
9c	Subjective predictor assessment	Predictors consisted of objective laboratory measurements and did not require subjective study-level assessment.	Not applicable
10	Sample size	The analysis used the complete available analytical dataset of 335 individuals; no additional sampling was performed.	Methods; Results
11	Missing data	Missingness occurred only for INR; fold-wise median imputation was used for machine-learning analyses, whereas individual INR analyses used available cases without imputation.	Methods: Data preprocessing; individual biomarker analysis
12a	Use/partition of data	Internal validation used 10×5 repeated stratified cross-validation, with preprocessing performed within training folds.	Methods: Machine learning model development and internal validation
12b	Predictor handling	Fold-wise standardization was used for scale-sensitive algorithms; transformations and standardization for conventional regression are specified.	Methods; Table 3
12c	Model building/ hyperparameters/ internal validation	Nine supervised algorithms with fixed parameter settings were evaluated without data-driven hyperparameter optimization. Full settings are provided in Supplementary Table S1.	Methods; Supplementary Table S1
12d	Cluster heterogeneity	The dataset was derived from a single centre; cluster-level heterogeneity analysis was therefore not applicable.	Not applicable
12e	Performance/ model comparison	ROC-AUC, classification metrics, calibration, Brier score, exploratory decision curve analysis, SHAP, and paired ROC-AUC comparisons were evaluated.	Methods
12f	Model updating	No external model updating or recalibration was performed.	Not applicable
12g	Predictions in external evaluation	No independent external evaluation dataset was available.	Not applicable
13	Class imbalance	Outcome classes were relatively balanced (175 vs 160); oversampling and synthetic data generation were not used.	Methods: Data preprocessing
14	Fairness	Fairness analyses could not be performed because relevant sociodemographic variables were unavailable.	Not applicable
15	Model output/ thresholds	The positive class and classification metrics are defined; a 0.50 threshold was used for the aggregate OOF confusion matrix, and biomarker cutoffs were derived from training folds using the Youden index.	Methods; Results
16	Development vs evaluation data	Model development and internal validation were performed within the same dataset using repeated cross-validation; no independent external dataset was used.	Methods; Discussion
17	Ethics	Ethics approval, approval number/date, Declaration of Helsinki compliance, and waiver of informed consent are reported.	Methods; Ethical considerations
18a	Funding	No external funding was received.	Funding
18b	Conflicts of interest	No financial or non-financial conflicts of interest were declared.	Conflict of interest
18c	Protocol	No separate prospective study protocol was registered for this retrospective analysis.	Not applicable
18d	Registration	The retrospective study was not prospectively registered.	Not applicable
18e	Data sharing	Data are available from the corresponding author upon reasonable request.	Data availability
18f	Code sharing	Analytical code was not deposited in a public repository.	Not publicly available

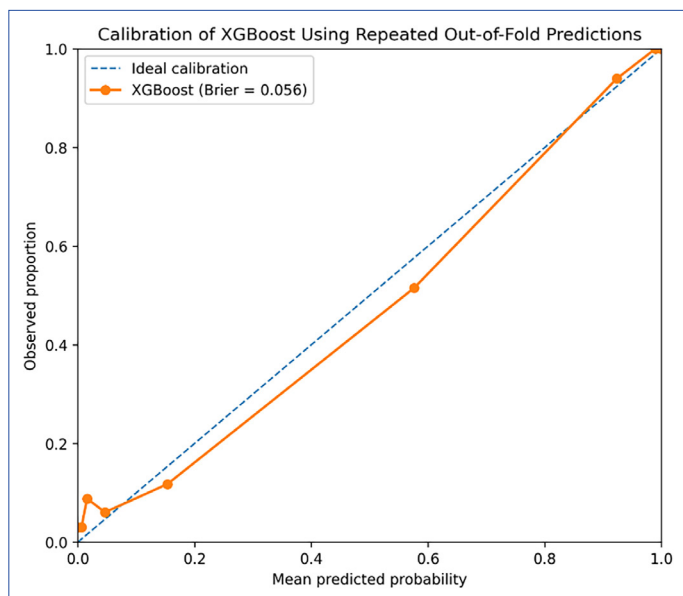
Appendix 2. Cont.

Item	Topic	How addressed in this study	Reported on page/section
19	Patient/public involvement	Patients or members of the public were not involved in the design, conduct, reporting, or dissemination planning of this retrospective analysis.	Not applicable
20a	Participant flow	The final analytical sample and outcome-group counts are reported. No longitudinal follow-up was involved.	Results: Study population and dataset characteristics
20b	Participant characteristics	Available biochemical characteristics are presented by outcome group; additional demographic and clinical characteristics were unavailable.	Table 1; Discussion: Limitations
20c	Comparison with development data	No separate external evaluation dataset was used.	Not applicable
21	Numbers in analyses	Overall sample size and group counts are reported; the INR complete-case sensitivity analysis reports the corresponding analytical sample.	Results
22	Full model specification	Model-specific estimator settings and preprocessing procedures are provided in Supplementary Table S1.	Supplementary Table S1
23a	Model performance	Repeated-CV discrimination and classification metrics, ROC-AUC uncertainty, calibration, exploratory decision analysis, and explainability results are reported.	Results; Table 4; Figures; Supplementary Figures
23b	Cluster heterogeneity	No multicentre or clustered evaluation was performed.	Not applicable
24	Model updating	No model updating following external evaluation was performed.	Not applicable
25	Interpretation	Findings are interpreted as internal classification performance rather than independent clinical diagnostic validity; incorporation bias and generalizability are discussed.	Discussion
26	Limitations	Outcome definition, residual incorporation bias, sample size, single-centre design, missing clinical information, lack of external validation, and calibration limitations are discussed.	Discussion: Limitations
27a	Unavailable/poor-quality inputs	INR missingness and its handling are described; the model is not proposed for clinical deployment.	Methods; Discussion
27b	User interaction / expertise	Clinical implementation and user requirements were not defined because the model is exploratory and not proposed for deployment.	Not applicable
27c	Future research / generalizability	External validation in larger multicentre cohorts with independently adjudicated outcomes is recommended before clinical application.	Discussion; Conclusions

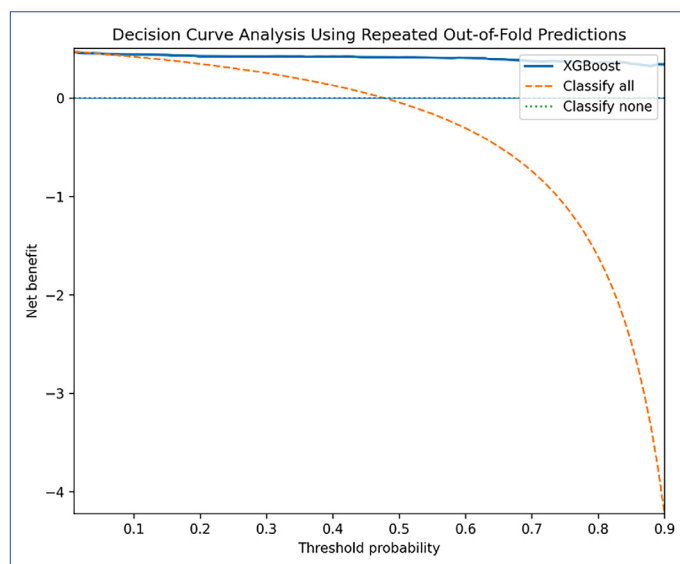
Checklist prepared against the 27-item TRIPOD+AI framework. Item descriptions are abbreviated. Page/section locations refer to the revised manuscript structure; final page numbers should be updated after the tracked-changes version is paginated.



Appendix 3. Aggregate out-of-fold confusion matrix of the XGBoost model based on 10×5 repeated stratified cross-validation. For each individual, predicted probabilities obtained across the 10 repetitions were averaged and converted to class predictions using a threshold of 0.50. Cell values show counts and row percentages.



Appendix 4. Calibration curve of the XGBoost model based on averaged repeated out-of-fold predictions from 10×5 repeated stratified cross-validation. Predicted probabilities obtained for each individual across the 10 repetitions were averaged before calibration assessment. The diagonal line represents perfect calibration. The aggregate out-of-fold Brier score was 0.056.



Appendix 5. Decision curve analysis of the XGBoost model based on averaged repeated out-of-fold predictions from 10×5 repeated stratified cross-validation. Net benefit is shown across threshold probabilities for the XGBoost model and the classify-all and classify-none strategies. The analysis was considered exploratory because the outcome was ICD-defined and no independent clinical endpoint was available.



Research Article

Specimen type as the dominant predictor of pre-analytical rejection in the emergency department: An explainable machine-learning (SHAP) analysis

Yunus Goren¹, Hakan Ozerol²

¹Department of Medical Biochemistry, Gaziantep City Hospital, Gaziantep, Türkiye

²Department of Emergency Medicine, Gaziantep City Hospital, Gaziantep, Türkiye

Abstract

Objectives: Preanalytical specimen rejection is a leading source of laboratory errors in emergency departments, but most evidence comes from small studies lacking a true denominator. We examined factors associated with specimen rejection in a complete 34-month cohort and assessed the extent to which apparent predictability was attributable to specimen type.

Methods: All tubes collected in the emergency department of a tertiary hospital between October 2023 and August 2026 were included. The outcome was tube rejection. Predictors were restricted to information available at the time of request: patient characteristics, timing (including hours since shift handover), workload, collection context, specimen type, and panel size. Four algorithms were compared using temporal validation on tubes collected between January and August 2026 and evaluated for discrimination, calibration, net benefit, subgroup performance, and Shapley additive explanations. Adjusted odds ratios (ORs) were calculated using standard errors clustered by presentation.

Results: Of 1,190,944 tubes collected during 404,247 presentations, 37,633 (3.16%) were rejected, with hemolysis accounting for 67.6% of rejections. LightGBM achieved an out-of-time area under the receiver operating characteristic curve (AUROC) of 0.780 (95% confidence interval [CI], 0.775–0.785), with a calibration slope of 0.983. Biochemistry tubes had an adjusted OR of 10.48 (95% CI, 10.05–10.93). Specimen type alone achieved an AUROC of 0.709, whereas all patient characteristics, timing, and workload variables combined achieved an AUROC of 0.601. The odds of rejection were highest during the final 4 hours of the 24-hour shift (OR, 1.15; 95% CI, 1.09–1.21). The lower unadjusted odds observed in pediatric and obstetric units were no longer evident after adjustment.

Conclusion: Specimen type was the strongest predictor of rejection. Models developed using laboratory information system data reflect differences in specimen mix and should be benchmarked against specimen type.

Keywords: Hemolysis, hospital emergency service, machine learning, pre-analytical phase, quality indicators, specimen handling

How to cite this article: Goren Y, Ozerol H. Specimen type as the dominant predictor of pre-analytical rejection in the emergency department: An explainable machine-learning (SHAP) analysis. Int J Med Biochem 2026;9(4):251–263.

Preanalytical errors account for most errors in the total testing process, and specimen rejection is their most visible manifestation [1–3]. Emergency departments are the highest-risk settings for specimen collection: sampling occurs under time pressure, frequently through intravenous catheters, and is performed by staff whose primary responsibility is not

phlebotomy. Reported rejection and hemolysis rates range from less than 1% of specimens in laboratory-wide series to more than 10% in emergency departments, with hemolysis being the leading cause in nearly every published series [4–7]. Two limitations recur in this literature. First, most studies count rejected specimens without a reliable denominator that

Address for correspondence: Yunus Goren, MD. Department of Medical Biochemistry, Gaziantep City Hospital, Gaziantep, Türkiye

Phone: +90 506 420 91 28 **E-mail:** ysgoren@gmail.com **ORCID:** 0000-0003-0344-2456

Submitted: September 23, 2026 **Revised:** October 07, 2026 **Accepted:** October 09, 2026 **Available Online:** October 15, 2026

OPEN ACCESS This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).



includes accepted specimens, so the reported rate depends on how the denominator was reconstructed. Second, cohorts are typically small, ranging from thousands to a few tens of thousands of specimens, leaving them underpowered to test the operational hypotheses they most frequently advance: that rejection rates increase at night, with hours on duty, and during periods of high departmental workload. When such associations have been reported, they have rarely been adjusted for the composition of the specimens collected, even though hemolysis affects serum biochemistry tubes far more frequently than hematology or blood gas specimens [4].

Machine learning has recently been applied to preanalytical error detection [8–12], with studies often reporting very high discriminatory performance. However, the variables responsible for this performance are seldom explicitly identified. If a model includes variables encoding specimen type, much of its apparent predictive ability may reflect the near-deterministic relationship between tube type and hemolysis rather than any patient- or process-level insights.

We assembled the complete emergency department laboratory records of a tertiary hospital over 34 consecutive months, including every accepted and rejected tube, and addressed three questions. Which factors are associated with specimen rejection after adjusting for specimen type? Are circadian, shift, and workload factors associated with rejection, and is the cohort sufficiently large to exclude associations of practical relevance? How much of an explainable machine learning model's discriminatory performance reflects specimen mix rather than modifiable process factors?

Materials and Methods

Setting and design

This retrospective cohort study included all specimen tubes collected in the emergency department of a tertiary care hospital between 6 October 2023 and 10 August 2026. The department comprises adult, obstetric, and four pediatric units. Patients are allocated to red, yellow, or green triage areas, and nursing staff work 24-hour shifts with a single handover at 08:00. Specimens are sent to a co-located core laboratory and screened upon receipt.

The study is reported in accordance with TRIPOD+AI [13]. This study was approved by the Gaziantep City Hospital Institutional Ethics Committee (Date: 22 July 2026, Decision No: 624/2026), which waived the requirement for informed consent for the secondary use of anonymized routine data. Approval details are provided on the title page.

The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

Data source and dataset construction

Records were extracted from the laboratory information system (LIS) at the individual test-request level, comprising 10,283,214 rows. Each row contained an anonymized pre-

sentation identifier, collection barcode, patient age and sex, emergency department unit, triage area, admission status, requested test, request timestamp, and, where applicable, rejection reason and laboratory timestamps.

Because rejection applies to a tube rather than an individual test, rows were aggregated by barcode, yielding 1,190,944 tubes. For each tube, we recorded the set of distinct tests requested, from which specimen type was derived and classified into one of six mutually exclusive categories: biochemistry (serum), complete blood count (CBC), blood gas, urine, coagulation, and other. Panel size was defined as the number of distinct tests requested for each tube.

Outcome

The primary outcome was tube rejection by the laboratory, defined as the presence of any rejection reason in any constituent test row. This definition was chosen because it corresponds to the operational event relevant to the emergency department, namely specimen loss and the need for repeat collection, as well as the rejection-rate quality indicator reported to laboratory management and harmonized internationally [14]. Rejection reasons were retained verbatim and grouped post hoc into specimen-quality reasons (hemolysis, clotting, insufficient volume, lipemia, icterus, inappropriate container or fill level, unseparated serum, temperature deviation, and breakage) and administrative reasons (absent blood-draw record, barcode and data-entry errors, duplicate entry, requested repeat sampling, and tests not performed in the emergency laboratory). Because these two groups reflect different processes, two secondary outcomes were analyzed: specimen-quality rejection, for which tubes rejected solely for administrative reasons were treated as non-events, and hemolysis-specific rejection.

Predictors

Predictors were restricted to information available at the time the request was entered. Patient variables included age, sex, emergency department unit, triage area, and admission status. Age was exported from the LIS in completed years; therefore, a value of 0 denoted an infant younger than 1 year and was analyzed as a valid value rather than as a missing or default entry. Timing variables included hour of day (sine and cosine terms), hours since the 08:00 handover, day of week, weekend, season, and calendar time. Workload variables included the number of tubes drawn in the same unit during the same hour and across the department during the same hour and day. Collection-context variables included the number of tubes drawn from the patient during the same minute and the interval since the first specimen collected during the presentation. Specimen variables included specimen type and panel size.

Triage area was missing for 1,278 tubes (0.107%); these tubes were retained in all analyses, and no imputation was performed. All other predictors were complete.

Screening for information leakage

Three candidate variables that encoded the outcome were excluded before model fitting.

The number of result rows per tube was excluded because accepted tubes accumulate repeat and reflex measurement rows, whereas rejected tubes do not (among CBC tubes, 97.4% of accepted specimens but only 4.2% of rejected specimens contained more than one row per test). This variable was replaced by the count of distinct requested tests.

Presentation-level totals (tubes, tests, and duration per presentation) were excluded because rejection triggers a repeat draw and therefore increases these totals, making them consequences of the outcome.

The rank of a tube within a presentation was excluded. For the 85% of tubes recorded in the same minute as another tube from the same patient, this rank was derived from the row order in the data export rather than from an observed sequence, and it encoded specimen type: 49% of tubes ranked first were biochemistry tubes, compared with 3.5% of those ranked fifth. Retaining this variable increased the apparent area under the receiver operating characteristic curve (AUROC) by 0.075 while reducing within-specimen-type discrimination below chance levels.

LIS timestamps were used only to describe turnaround times and were never included as predictors in any model.

Statistical analysis

The cohort was divided temporally. Tubes collected from October 2023 through December 2025 formed the development set ($n=910,776$), and tubes collected in 2026 formed the test set ($n=280,168$). Within the development set, the final 10% of observations by time were held out for early stopping.

Four algorithms were fitted: LightGBM [15], XGBoost [16], random forest, and L2-penalized logistic regression with a restricted cubic spline for age. No class weighting was applied. Gradient-boosting hyperparameters were explored across twelve configurations (15–255 leaves; learning rates of 0.015–0.05), and the number of boosting rounds was determined by early stopping using the time-ordered holdout set described above. Discrimination was summarized using AUROC and the area under the precision–recall curve (AUPRC), and calibration was assessed using the calibration slope, calibration-in-the-large (CITL), and scaled Brier score [17]. AUROC confidence intervals were calculated using bootstrap resampling clustered by presentation (400 replicates). Paired differences between algorithms were tested using the DeLong method [18]. Clinical usefulness was assessed using decision-curve analysis [19] and the proportion of rejections captured when a specified fraction of the highest-risk tubes was flagged.

To determine whether discrimination was limited by modeling or by available information, the same twelve gradient-boosting configurations were applied to the 23-variable set and an expanded 85-variable set that included panel-group indica-

tors. Three additional blocks were then added sequentially: indicators for the 48 most frequently requested tests, rejection rates in the same unit over the preceding 7 and 30 days, and an indicator of prior rejection during the same presentation that had already been recorded at the time of the request. For this analysis, the highest test-set AUROC across configurations is reported. This represents an optimistic upper bound and is therefore conservative with respect to the conclusion that discrimination is limited by available information.

Adjusted associations were estimated using logistic regression with standard errors clustered by presentation. Categorical predictors were entered as indicator variables, with the reference categories specified in Table 1. Age was modeled using linear and quadratic terms, and workload was entered as a standardized variable. Predictor blocks were added sequentially and evaluated in the same test set to quantify the contribution of each block.

Model behavior was interpreted using Shapley additive explanations (SHAP), [20, 21] calculated with the TreeSHAP algorithm implemented in the shap library on 20,000 randomly selected test-set tubes. The tree-path-dependent formulation was used, in which the contributions of absent predictors were estimated from the training-data coverage of each tree path; therefore, no separate background dataset was required. Values are expressed on the model's log-odds scale. SHAP values describe how the fitted model uses each predictor rather than its causal or biological importance [22], and correlated predictors, such as specimen type and panel size, share attribution according to the structure learned by the trees [23].

For this revision, the complete pipeline was rerun from the raw LIS export using a fixed random seed (2026), and all settings were recorded. In the rerun, specimen type was assigned based on the defining test for each tube. As a result, 38,822 tubes (3.3%) originally classified as other specimens were reclassified into the blood-gas and coagulation categories. The rerun was used to assess the robustness of the main findings and to perform the specimen-quality analysis. All other estimates were derived from the original analysis, and the software versions listed below correspond to those used in the original analysis.

All analyses were performed in Python 3.11.15 (Python Software Foundation, Wilmington, Delaware, USA) with the open-source libraries scikit-learn 1.8.0, statsmodels 0.14.6, XGBoost 3.2.0, shap 0.51.0 and LightGBM 4.7.0 (Microsoft Corp., Redmond, Washington, USA). A two-sided $p<0.05$ was considered statistically significant.

Sample size

No sample size was specified in advance because the study included every eligible tube collected during the extraction period. Sample size adequacy was assessed retrospectively according to the criteria of Riley et al. [24] for prediction models with a binary outcome. Assuming a development-set prevalence of 3.02%, a Cox–Snell R^2 of 0.034

Table 1. Adjusted odds ratios for specimen rejection (n=1,190,944 tubes), from logistic regression with standard errors clustered on presentation

Variable	Adjusted OR	95% CI	p
Specimen type: biochemistry (serum)	10.48	10.05–10.93	<0.001
Specimen type: coagulation	2.41	2.24–2.61	<0.001
Specimen type: urine	1.89	1.78–2.02	<0.001
Triage area: red	1.75	1.66–1.86	<0.001
Specimen type: complete blood count	1.34	1.28–1.41	<0.001
Age, per 10 years	1.17	1.14–1.21	<0.001
Triage area: yellow	1.14	1.08–1.19	<0.001
Inpatient (vs outpatient)	1.12	1.08–1.16	<0.001
Male sex	1.12	1.09–1.15	<0.001
Specimen type: blood gas	1.11	1.03–1.18	0.003
Last 4 h of the 24-h nursing shift (binary indicator)	1.07	1.01–1.14	0.03
Tubes drawn simultaneously, per tube	1.07	1.06–1.08	<0.001
Concurrent workload, per SD	1.04	1.02–1.05	<0.001
Hours since handover, per SD	1.02	0.99–1.06	0.15
Pediatric unit	1.02	0.95–1.09	0.62
Weekend	1.00	0.97–1.02	0.88
Age, quadratic (plateau) term	0.993	0.990–0.996	<0.001
Obstetric unit	0.99	0.92–1.06	0.71
Night, 00:00–08:00	0.97	0.89–1.05	0.46
Evening, 16:00–24:00	0.95	0.91–1.00	0.05

Reference categories: other specimen type; green triage area; female sex; outpatient; adult unit; weekday; daytime (08:00–16:00). Before adjustment for specimen type the pediatric unit OR was 0.82 and the obstetric unit OR 0.64. The last-4-h term is a binary indicator entered in this model; Appendix 4 reports the same exposure as six 4-h blocks in a separate model, in which the final block gave an OR of 1.15 (1.09–1.21). CI: Confidence interval; OR: Odds ratio; SD: Standard deviation.

(corresponding to the observed C-statistic of 0.78), and up to 150 candidate predictor parameters, at least 39,168 development tubes (1,182 events) would be required to maintain expected shrinkage at or above 0.90 and minimize optimism in apparent model fit. The development set contained 910,776 tubes and 27,489 events, approximately 23 times the required sample size. The 2026 test set contained 10,144 events, substantially exceeding the recommended minimum of 100–200 events for validation [25].

Results

Cohort

Of the 1,190,944 tubes collected during 404,247 presentations, 37,633 (3.16%) were rejected (Table 2). Rejected and accepted tubes differed most in specimen type (standardized mean difference [SMD] for biochemistry, +1.08), age (+0.33), unit (adult, +0.31), triage area (red, +0.23), and concurrent workload (+0.19), whereas they were essentially identical in hour of request (−0.01), shift (≤ 0.02), and weekend status (0.00).

A black-area code, which was not part of the triage scheme used in this setting, appeared on 2,324 tubes (0.195%), almost all from the adult unit and distributed evenly throughout the study period, consistent with a data-entry origin. These tubes were analyzed as recorded. Of the 40,644 tubes recorded for patients aged 0 years, 99.6% came from pediatric units, consistent with actual infant age rather than a default entry.

Causes of rejection and an artifactual trend

Hemolysis accounted for 25,439 rejections (67.6%), clotting for 4,518 (12.0%), insufficient volume for 2,487 (6.6%), absent blood-draw records for 2,130 (5.7%), and lipemia for 720 (1.9%) (Appendix 1a).

The category of absent blood-draw records first appeared in 2025 (n=390) and increased to 1,740 in 2026, accounting for 17.2% of that year's rejections. After excluding administrative categories, the rejection rate remained stable throughout the study period: 2.70%, 3.08%, 2.74%, and 2.90% for 2023–2026, compared with 2.78%, 3.18%, 2.92%, and 3.62% for all rejections (Appendix 1b, c). Administrative reasons therefore contributed 0.08, 0.10, and 0.18 percentage points to the annual rejection rate in 2023–2025, respectively, and 0.72 percentage points in 2026. The apparent 24% relative increase between 2025 and 2026 (chi-square test, $p=1.3 \times 10^{-62}$) was therefore largely a coding artifact. Over the same period, the proportion of rejections attributed to clotting decreased from 18.8% to 7.0%, whereas the proportion attributed to hemolysis remained between 63% and 70%.

Model performance

LightGBM achieved an out-of-time AUROC of 0.780 (95% CI, 0.775–0.785), AUPRC 0.146 against a base rate of 3.62%, scaled Brier +5.66%, calibration slope 0.983 and CITL 0.239 (Appendix 2; Fig. 1a–c). Random forest reached 0.776, XGBoost 0.774

Table 2. Characteristics of 1,190,944 emergency department specimen tubes by rejection status, with standardized mean differences

Characteristic	All tubes	Accepted	Rejected	SMD
Tubes, n	1,190,944	1,153,311	37,633	
ED presentations, n	404,247	403,822	34,373	
Age, mean (SD)	39.2 (25.7)	39.0 (25.7)	47.3 (24.8)	+0.331
Age, median (IQR)	38 (18–61)	38 (17–60)	49 (27–68)	
0–1 years, n (%)	70,636 (5.9)	69,164 (6.0)	1,472 (3.9)	-0.096
2–17, n (%)	224,036 (18.8)	220,086 (19.1)	3,950 (10.5)	-0.244
18–44, n (%)	388,807 (32.6)	377,602 (32.7)	11,205 (29.8)	-0.064
45–64, n (%)	259,011 (21.7)	249,316 (21.6)	9,695 (25.8)	+0.098
65–79, n (%)	175,310 (14.7)	167,495 (14.5)	7,815 (20.8)	+0.164
≥80, n (%)	73,144 (6.1)	69,648 (6.0)	3,496 (9.3)	+0.122
Sex, n (%)				
Female	659,954 (55.4)	640,344 (55.5)	19,610 (52.1)	-0.069
Male	530,990 (44.6)	512,967 (44.5)	18,023 (47.9)	+0.069
Emergency department unit, n (%)				
Adult	827,370 (69.5)	796,506 (69.1)	30,864 (82.0)	+0.305
Pediatric 1	226,363 (19.0)	222,217 (19.3)	4,146 (11.0)	-0.232
Obstetric	76,301 (6.4)	74,728 (6.5)	1,573 (4.2)	-0.103
Pediatric 2	28,655 (2.4)	28,174 (2.4)	481 (1.3)	-0.086
Pediatric 3	26,448 (2.2)	25,945 (2.2)	503 (1.3)	-0.069
Pediatric 4	5,807 (0.5)	5,741 (0.5)	66 (0.2)	-0.056
Triage area, n (%)				
Red area	190,473 (16.0)	181,184 (15.7)	9,289 (24.7)	+0.225
Yellow area	880,019 (73.9)	853,897 (74.0)	26,122 (69.4)	-0.103
Green area	116,850 (9.8)	114,757 (10.0)	2,093 (5.6)	-0.165
Black area	2,324 (0.2)	2,195 (0.2)	129 (0.3)	+0.030
Missing	1,278 (0.1)	1,278 (0.1)	0 (0.0)	-0.047
Specimen type, n (%)				
Complete blood count	335,042 (28.1)	331,031 (28.7)	4,011 (10.7)	-0.466
Biochemistry (serum)	309,048 (25.9)	281,953 (24.4)	27,095 (72.0)	+1.082
Other	279,155 (23.4)	276,549 (24.0)	2,606 (6.9)	-0.486
Blood gas	118,946 (10.0)	117,574 (10.2)	1,372 (3.6)	-0.260
Urine	111,092 (9.3)	109,494 (9.5)	1,598 (4.2)	-0.209
Coagulation	37,661 (3.2)	36,710 (3.2)	951 (2.5)	-0.039
Inpatient, n (%)	203,758 (17.1)	195,522 (17.0)	8,236 (21.9)	+0.125
Weekend, n (%)	338,601 (28.4)	327,975 (28.4)	10,626 (28.2)	-0.004
Timing and workload				
Hour of request, median (IQR)	15 (10–19)	15 (10–19)	14 (10–19)	-0.014
Hours since the 08:00 handover, median (IQR)	10 (5–14)	10 (5–14)	10 (5–14)	-0.007
Workload: tubes in unit that hour, median (IQR)	34 (17–49)	34 (17–49)	38 (23–52)	+0.187
Tubes drawn simultaneously, median (IQR)	3 (2–4)	3 (2–4)	3 (3–4)	+0.170
Shift 08–16, n (%)	442,919 (37.2)	428,566 (37.2)	14,353 (38.1)	+0.020
Shift 16–24, n (%)	542,126 (45.5)	525,428 (45.6)	16,698 (44.4)	-0.024
Shift 24–08, n (%)	205,899 (17.3)	199,317 (17.3)	6,582 (17.5)	+0.005
Study period, n (%)				
Development (Oct 2023–Dec 2025)	910,776 (76.5)	883,287 (76.6)	27,489 (73.0)	-0.082
Test (Jan–Aug 2026)	280,168 (23.5)	270,024 (23.4)	10,144 (27.0)	+0.082
Missing data				
Triage area missing, n (%)	1,278 (0.107)			
Age/sex/unit missing, n	0			

Values are N (%) unless stated otherwise. Indented rows are categories of the preceding variable. Hours are relative to the 08:00 nursing handover. The black-area triage code is not part of the triage scheme in use in this department and is considered a data-entry artefact; tubes were analyzed as recorded. IQR: Interquartile range; SD: Standard deviation; SMD: Standardized mean difference.

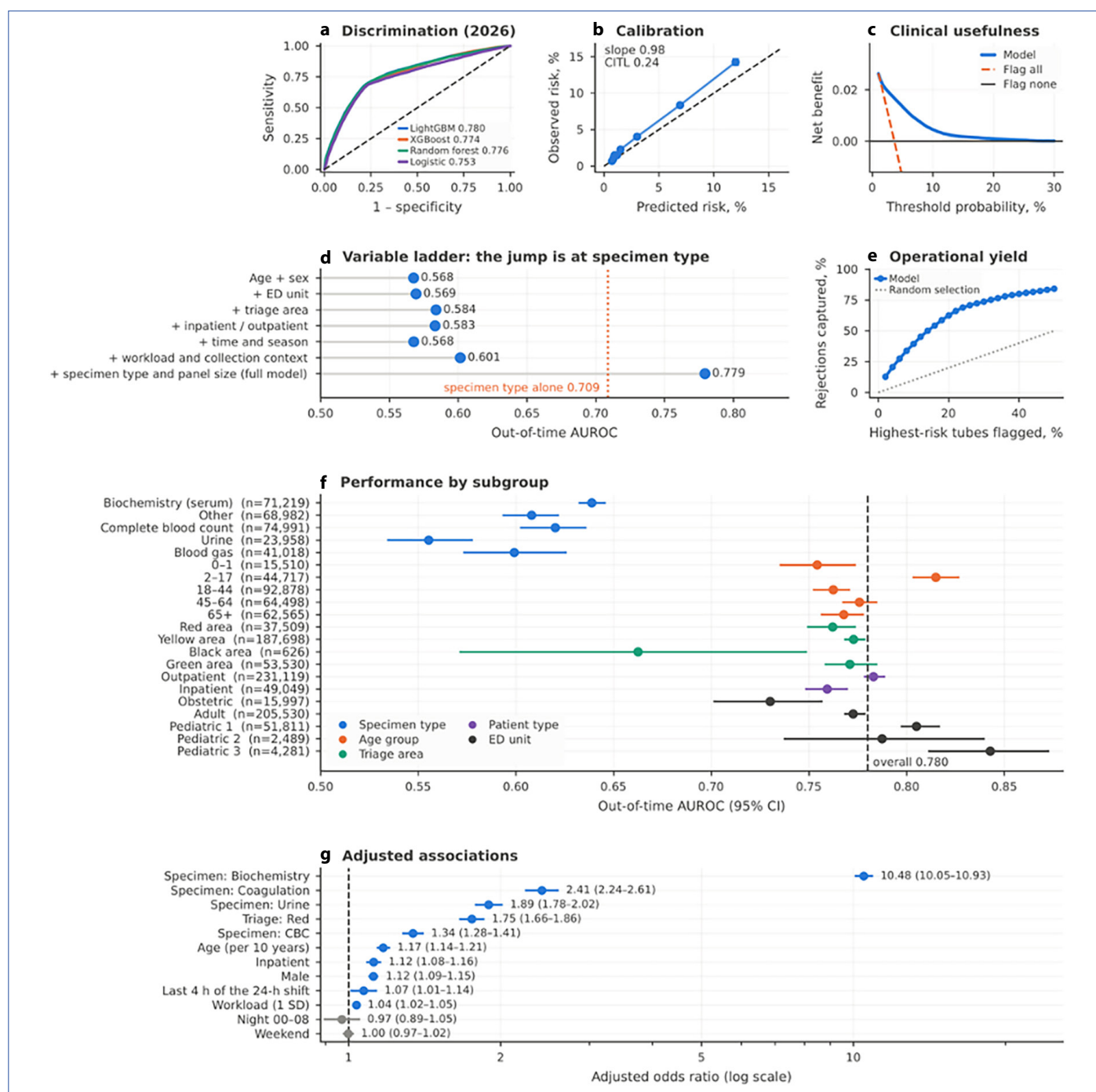


Figure 1. Model performance and predictors of specimen rejection. (a) ROC curves for four algorithms in the 2026 out-of-time test set. (b) Calibration of the LightGBM model by decile of predicted risk, with Wilson CIs; the dashed line indicates perfect calibration. (c) Decision-curve analysis comparing the model with flag-all and flag-none strategies. (d) Out-of-time AUROC as predictor blocks are added successively; the dotted vertical line marks the AUROC of a model containing specimen type alone. (e) Proportion of rejections captured as a function of the proportion of highest-risk tubes flagged; the diagonal represents random selection. (f) Out-of-time AUROC with 95% CIs within subgroups defined by specimen type, age group, triage area, patient type and emergency department unit; the dashed vertical line indicates overall performance. (g) Adjusted ORs with 95% CIs; gray markers indicate intervals crossing unity.

CI: Confidence interval; OR: Odds ratio; SD: Standard deviation.

and penalized logistic regression 0.753. Paired DeLong differences favored LightGBM over all three comparators (all $p < 0.001$), but differences between tree-based algorithms

were small (≤ 0.007) and their ranking was not stable: in the re-run described below, random forest reached 0.787 and XGBoost 0.783, against 0.783 for LightGBM.

Calibration was achieved without class weighting. With class weighting, discrimination was unchanged but the calibration slope rose to 1.27 and CITL fell to -3.05 .

Where the discrimination originates

Adding predictor blocks successively (Fig. 1d) gave 0.568 for age and sex, 0.569 after unit, 0.584 after triage area and 0.583 after admission status; the timing block did not improve discrimination (0.568), workload and collection context raised it to 0.601, and specimen type with panel size raised it to 0.779. Specimen type alone, a single six-level variable, achieved 0.709.

Is 0.78 a modeling limit or an information limit?

In the original analysis, no configuration and no feature set exceeded an out-of-time AUROC of 0.781 (Appendix 3). The 85-variable model reached 0.781 and the 23-variable model 0.780, and every added feature block reduced discrimination (to 0.775–0.777). The ceiling therefore reflects the information content of the recorded variables, not the choice of algorithm or hyperparameters.

Adjusted associations

Specimen type was the strongest predictor (Table 1; Fig. 1g). Relative to the "other" specimen category, biochemistry tubes carried an adjusted OR of 10.48 (10.05–10.93), coagulation tubes 2.41 (2.24–2.61), urine 1.89 (1.78–2.02), CBC 1.34 (1.28–1.41) and blood gas 1.11 (1.03–1.18).

Among patient variables, red-area triage carried 1.75 (1.66–1.86) and yellow-area 1.14 (1.08–1.19) relative to green; age 1.17 (1.14–1.21) per 10 years, with a plateau term of 0.993 (0.990–0.996); male sex and admission 1.12 each.

Two associations observed in unadjusted analysis disappeared after adjustment for specimen type. The pediatric unit OR moved from 0.82 to 1.02 (0.95–1.09; $p=0.62$) and the obstetric unit OR from 0.64 to 0.99 (0.92–1.06; $p=0.71$). The lower crude rejection rates in pediatric and obstetric units were therefore largely accounted for by the different mix of specimens requested there, and the adjusted estimates were close to unity.

Time of day, shift hours and workload

Because nursing shifts are 24 h long with a single 08:00 handover, hours on duty and clock time are the same variable and cannot be separated in this design.

With hours since handover entered in 4-h blocks and the first as reference (Appendix 4, 5a), the OR was at or near unity through hours 4–16, rose to 1.05 (1.01–1.10) for hours 16–20 and reached 1.15 (1.09–1.21) in the final 4 h (04:00–08:00). Restricted to biochemistry tubes with hemolysis as the cause, the final-block OR was 1.14 (1.07–1.22). The crude hemolysis rate in biochemistry tubes rose from 7.4% at handover to 9.2%–9.7% in the last three hours of the shift.

Weekend status was not associated with rejection (OR, 1.00; 0.97–1.02; $p=0.88$). Concurrent workload reached statistical significance but not clinical relevance (OR, 1.04 per SD; 1.02–1.05).

Interpretation

SHAP attribution ranked panel size first (mean |SHAP|, 0.571), followed by specimen type (0.274), calendar time (0.190), age (0.136) and triage area (0.118); all remaining predictors contributed ≤ 0.057 (Fig. 2a, b). These values describe how much each variable moved the model's predictions, not biological importance. Because large panels are ordered almost exclusively on serum biochemistry tubes, panel size and specimen type share credit, and the high rank of panel size indicates that it carries much of the specimen-related information used by the model, not that ordering more tests causes rejection.

Dependence plots (Fig. 2c) showed a monotonic age relationship, with markedly lower predicted risk below 5 years, a steep rise to age 18 and a plateau after 60; a step-like panel-size pattern, with lower predicted risk for single-test tubes and higher predicted risk for panels of 15 or more tests (up to +1.2 log-odds); and an ordering of triage areas from red to green.

Subgroup performance

Discrimination was preserved across every patient stratum (Appendix 6; Fig. 1f), from 0.730 (0.701–0.757) in the obstetric unit to 0.843 (0.811–0.873) in one pediatric unit, with 0.754–0.815 across age groups and 0.762–0.773 across the red, yellow and green triage areas; the small black-area stratum ($N = 626$) gave 0.663 (0.571–0.749).

Within strata defined by specimen type, discrimination fell as expected when the dominant predictor is held constant: 0.639 for biochemistry, 0.620 for CBC, 0.608 for other specimens, 0.599 for blood gas and 0.555 for urine. All remained above chance, although calibration slopes varied between strata (0.84–1.42).

Restricting the outcome to hemolysis

Hemolysis was almost exclusively a biochemistry-tube phenomenon: 24,904 of 25,439 hemolysis rejections (97.9%) occurred in biochemistry tubes, where hemolysis represented 91.9% of all rejections. Hemolysis was recorded on 181 coagulation tubes and, implausibly, on 21 specimens in which serum hemolysis is not assessed; these were treated as data-entry errors (Fig. 3a).

Restricting the outcome to hemolysis while retaining all tubes raised the out-of-time AUROC to 0.907 (0.906–0.909). However, specimen type alone predicted this outcome at 0.871, and specimen type with panel size at 0.892 (Table 3; Fig. 3b). Models in this comparison were developed within the stated population and outcome, so their estimates differ slightly from the stratified evaluation of the primary model (Appendix 6). The increment attributable to every patient, timing, workload and collection variable combined was therefore 0.015.

Restricting both the population to biochemistry tubes and the outcome to hemolysis, the clinically coherent design, gave 0.652 (0.644–0.659) across 71,219 test-set tubes and 6,206 events, against 0.633 for any rejection in the same population. Within biochemistry tubes, age alone gave 0.585 and all patient variables together 0.609.

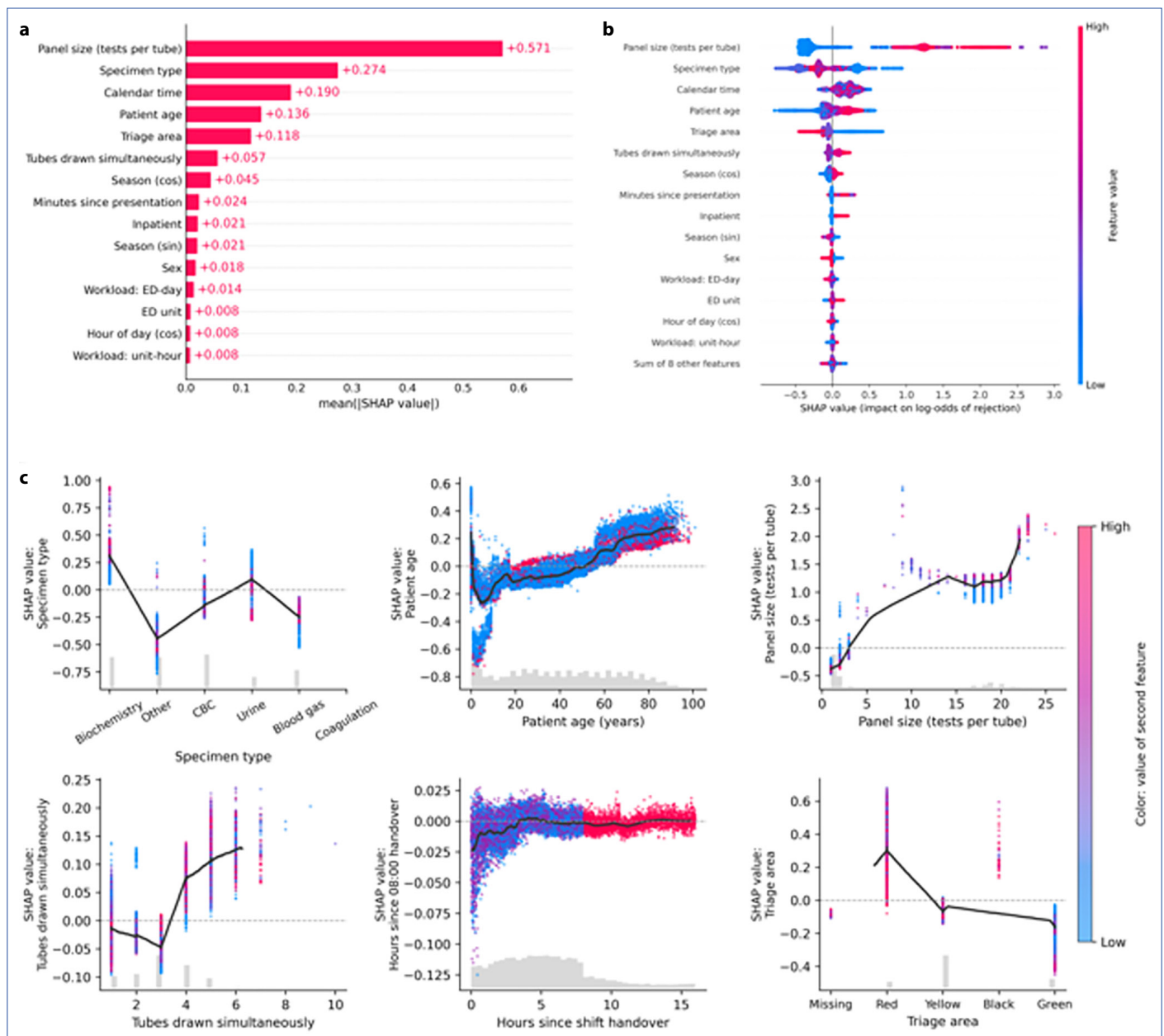


Figure 2. SHAP interpretation of the LightGBM model. (a) Mean absolute SHAP value for the 15 highest-ranked predictors. (b) Beeswarm plot of SHAP values for the 16 highest-ranked predictors; color encodes predictor value; positive values indicate higher predicted risk. (c) Dependence plots for specimen type, patient age, panel size, tubes drawn simultaneously, hours since the 08:00 handover and triage area; gray histograms show the marginal distribution and the black line a moving average. SHAP values are on the log-odds scale and describe the model's use of each predictor, not causal or biological importance; correlated predictors, such as panel size and specimen type, share attribution.

SHAP: Shapley additive explanations.

Re-run and specimen-quality outcome

The re-run from the raw export reproduced the cohort counts exactly (1,190,944 tubes; 37,633 rejections; 10,144 test-set events). LightGBM reached an out-of-time AUROC of 0.783 (0.777–0.787), the twelve configurations 0.776–0.784 and the four algorithms 0.766–0.787; the adjusted OR for the final 4 h of the shift was 1.15 (1.09–1.21) (Appendix 7, 8). With administrative-only rejections treated as non-events (34,428 events; 8,128 in the test set), discrimination was

0.838 (0.834–0.842) against 0.797 for specimen type alone, an increment similar to that for any rejection in the same run (0.783 against 0.737). Specimen type remained the strongest predictor (biochemistry OR, 20.9; 19.6–22.2), the pediatric and obstetric unit ORs moved from 0.80 and 0.58 before adjustment to 1.03 (0.96–1.10) and 0.95 (0.88–1.02) after it, and the final 4 h of the shift carried an OR of 1.11 (1.05–1.18). Concurrent workload was no longer associated with rejection (OR, 0.98 per SD; 0.97–1.00).

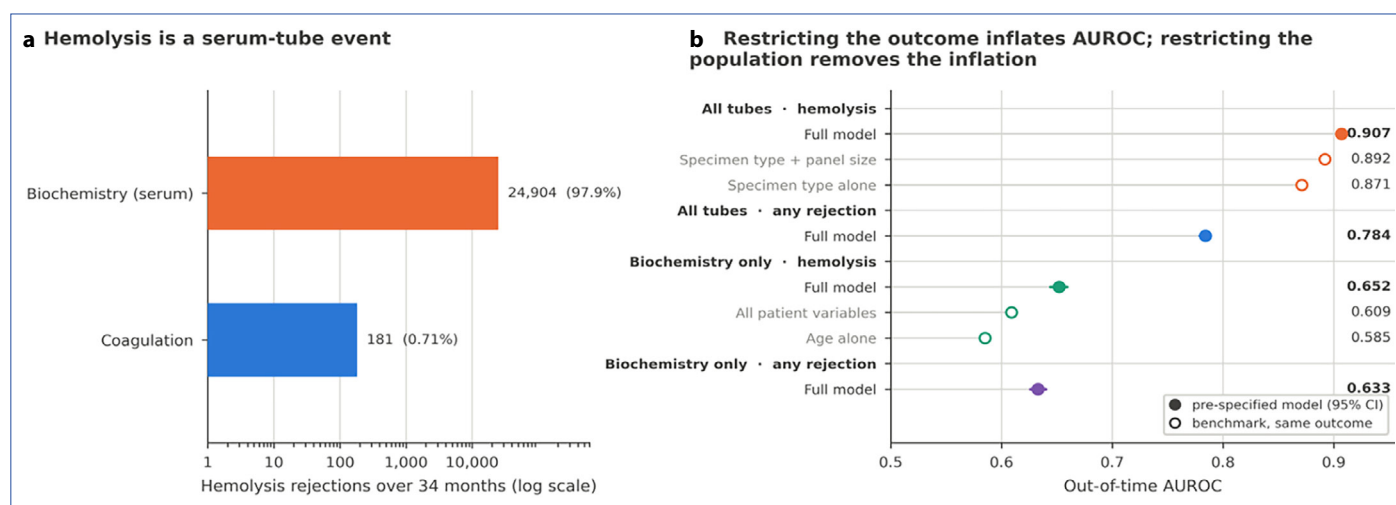


Figure 3. The specimen-type trap in hemolysis prediction. (a) Number of hemolysis rejections recorded over the 34-month cohort in the two specimen types in which serum hemolysis is assessed, on a logarithmic scale. A further 21 hemolysis codes recorded on urine, complete blood count and blood gas specimens were treated as data-entry errors and are not shown. (b) Out-of-time AUROC for each combination of analysis population and outcome. Filled markers are the pre-specified model with its 95% CI; open markers are benchmark models fitted to the same outcome. Restricting the outcome to hemolysis while retaining all tubes raises AUROC to 0.907, but specimen type entered alone reaches 0.871 for the same outcome; restricting the population to the specimen type in which the outcome can occur removes the inflation.

AUROC: Area under the receiver operating characteristic curve.

Clinical usefulness

Flagging the highest-risk 4% of tubes captured 20.5% of rejections at a positive predictive value (PPV) of 18.6% against a base rate of 3.62%; 10% captured 39.4% (PPV, 14.3%) and 20% captured 62.5% (PPV, 11.3%), a 3.1-fold enrichment. Decision-curve analysis showed net benefit above both flag-all and flag-none strategies at threshold probabilities of 1%–30% (Fig. 1c, 1e).

Turnaround

Median intervals from request were 53 min to result, 61 min to technician approval and 60 min to specialist approval; a rejection was recorded at a median of 62 min (Appendix 1d).

For a rejected specimen this interval is lost in full, and a repeat draw with a further cycle follows.

Discussion

In 1.19 million consecutive emergency department tubes, specimen type was the strongest predictor of laboratory rejection. Biochemistry serum tubes had tenfold higher adjusted odds of rejection than tubes in the "other" specimen category, and specimen type alone discriminated better (AUROC, 0.709) than all patient, timing and workload variables combined (0.601). Because the design is observational, these findings describe associations and sources of predictive discrimina-

Table 3. Effect of restricting the outcome to hemolysis and the population to biochemistry tubes

Population	Outcome	Test-set N	Events	Event rate, %	AUROC	95% CI
All tubes	Any rejection	280,168	10,144	3.62	0.784	0.779–0.788
All tubes	Hemolysis	280,168	6,463	2.31	0.907	0.906–0.909
All tubes - specimen type alone	Hemolysis	280,168	6,463	2.31	0.871	–
All tubes - specimen type + panel size	Hemolysis	280,168	6,463	2.31	0.892	–
Biochemistry (serum)	Any rejection	71,219	7,016	9.85	0.633	0.626–0.640
Biochemistry (serum)	Hemolysis	71,219	6,206	8.71	0.652	0.644–0.659
Biochemistry - age alone	Hemolysis	71,219	6,206	8.71	0.585	–
Biochemistry - all patient variables	Hemolysis	71,219	6,206	8.71	0.609	–
Complete blood count	Any rejection	74,991	1,037	1.38	0.598	0.580–0.614
Blood gas	Any rejection	41,018	261	0.64	0.610	0.582–0.643
Urine	Any rejection	23,958	639	2.67	0.520	0.498–0.542

Models in this table were developed and evaluated within the stated population and outcome. Across the full 34-month cohort, 24,904 of 25,439 hemolysis rejections (97.9%) occurred in biochemistry tubes and 181 in coagulation tubes; hemolysis was also coded on 21 specimens in which serum hemolysis is not assessed (12 urine, seven complete blood count, two blood gas), and these records were treated as data-entry errors. The all-tubes, any-rejection estimate in this table (0.784) therefore differs marginally from the primary model (0.780; 95% CI, 0.775–0.785; Appendix 2). AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval.

tion rather than causal effects; they nevertheless define what a preanalytical prediction model built on routine laboratory data can be expected to deliver.

The overall rejection rate of 3.16% and the hemolysis share of 67.6% are consistent with the published literature. Hemolysis accounts for 40%–70% of unsuitable specimens in clinical laboratories [4], and a recent analysis of approximately 11 million specimens attributed 69.6% of laboratory errors to hemolysis [3]. Laboratory-wide rejection rates are lower than ours: 0.65% in a Turkish clinical biochemistry laboratory that excluded emergency samples [7], and 0.79% in a recent Turkish series in which the emergency department nevertheless had the highest unit-level rate, 1.92% [26]. Hemolysis is considerably more frequent in emergency care: 6.8%–19.8% in most departments included in a systematic review and above 30% in two [5], and 17.2% of adult emergency encounters and 14.0% of pediatric emergency encounters with intravenous catheter placement in recent US cohorts [27, 28]. Rejection is less frequent than these prevalences because it is applied only when hemolysis threatens the analytes requested, so comparisons between studies must take this difference into account.

Specimen type is not a nuisance variable. It is measurable, known at order entry and biologically coherent: hemolysis interferes selectively with serum biochemistry [4]. Any model trained on LIS data will therefore encode specimen type, whether or not it is named, and reported discrimination should be benchmarked against a specimen-type-only model rather than against chance. In our data that benchmark was 0.709, so the full model's 0.780 represents an increment of 0.071 attributable to everything else recorded.

The hemolysis analysis makes this concrete. Restricting the outcome to hemolysis raises apparent discrimination to 0.907. Yet specimen type alone reaches 0.871 for the same outcome, because 97.9% of hemolysis occurs in biochemistry tubes: a model reported at 0.907 without that benchmark is, in effect, a serum-tube detector. Restricted to biochemistry tubes, where the question is clinically meaningful, discrimination falls to 0.652. Studies of hemolysis prediction should report the specimen-type-only comparator alongside their headline figure and restrict the analysis population to the specimen types in which the outcome can occur.

Published machine-learning applications in the preanalytical phase report much higher discrimination, but they address a different question with features that contain the error signal. Delta-check models for wrong-blood-in-tube errors reached AUROCs of 0.97–0.998 using consecutive results of multiple analytes, with errors created by simulated sample swaps [10, 11]. A model detecting falsely elevated point-of-care potassium in 3,489 emergency encounters reached 0.995 using a panel that included the potassium result itself [9], and image-based classification of serum hemolysis, icterus and lipemia reached 0.983–0.999 in 16,427 sample images [8]. These models detect an error after analysis or imaging, whereas ours estimates risk before collection, a harder task. In routine use, a wrong-blood-in-tube model developed on retrospective data

achieved a PPV of only 10.9% [12]. The PPVs of 11%–19% at our operational thresholds, and within-specimen-type AUROCs of 0.56–0.64, should be read in that light.

The main associations reported here, namely the dominance of specimen type, the disappearance of the pediatric and obstetric unit associations after adjustment and the late-shift excess, were estimated by logistic regression (Table 1) and do not depend on the machine-learning models; penalized logistic regression also reached an AUROC of 0.753, close to the 0.780 of LightGBM. This agrees with systematic evidence that machine learning offers no consistent advantage over logistic regression for clinical prediction [29]. The machine-learning framework was used for two narrower purposes: to establish an information ceiling, by showing that no configuration, feature set or algorithm moved discrimination materially beyond 0.78, and to characterize non-linear relationships, such as the age and panel-size profiles, without specifying their form in advance. It was not intended to deliver a deployable prediction tool.

Pediatric and obstetric units had lower unadjusted odds of rejection (ORs, 0.82 and 0.64), but both associations disappeared after adjustment for specimen type, because these units order proportionally fewer serum biochemistry panels. This does not contradict reports of frequent hemolysis in pediatric emergency care. In 10,462 pediatric emergency encounters with intravenous catheter placement, hemolysis affected 14.0% of encounters and 20.1% of those in infants, rates similar to the 17.2% reported in adult emergency encounters [27, 28]. Those studies measured hemolysis per encounter among patients who had blood drawn, whereas our denominator includes every tube ordered, with a specimen mix that differs between units. Unadjusted comparisons between collection areas, a common form of internal quality benchmarking, will therefore misattribute differences in specimen mix to collection technique.

Weekend status was not associated with rejection, and concurrent workload, although statistically significant at this sample size, corresponded to a 4% change in odds per SD. This association disappeared when administrative rejections were excluded. The closest published evidence, on blood-culture contamination, showed a modest and imprecise excess in the busiest quartile of hourly emergency department volume [30]. Odds were close to unity through the first 16 h of the 24-h shift, rose slightly at 16–20 h and were highest in the final 4 h (adjusted OR, 1.15). A recent Turkish series observed the same pattern by clock time: rejection was highest between 04:00 and 08:00 (incidence rate ratio, 1.63) [26]. In industrial studies, risk rises both on night shifts and with hours on duty, and the two have been described as separate components [31]; among hospital nurses, error risk increases with shifts exceeding 12 h [32]. In our single-handover roster the last hours of duty coincide with the early morning, and fatigue was not measured. The excess may therefore reflect accumulated fatigue, circadian factors, early-morning staffing or case mix, or a combination, and these explanations cannot be distinguished with the present data; designs with staggered handovers, or comparisons between rosters, would be required.

An administrative category silently inflated a quality indicator. The institutional rejection rate appeared to rise by 24% between 2025 and 2026. Most of that rise (0.54 of 0.70 percentage points) was attributable to administrative reasons, chiefly a category introduced in 2025 that records a missing blood-draw entry rather than a defective specimen. Once administrative categories were removed, the rate remained between 2.7% and 3.1% in every year. The numerator of this widely used quality indicator [14] is defined by whatever categories the LIS offers, so any institution reporting a trend should confirm that the category set was stable over the reporting window; ours was not, and the effect was large enough to have prompted an unnecessary intervention.

Leakage in routine laboratory data is subtle. Result-row counts and the export-derived position of a tube within a presentation would both be natural choices for an analyst, yet both encoded the outcome or specimen type. A variable that raises overall AUROC while driving within-stratum discrimination below chance, as tube position did, should prompt suspicion.

The ceiling is an information limit (Appendix 5b). Tuning configurations and feature sets extending to individual test flags, unit-level rejection history and prior rejections known at the time of request all converged on an AUROC of about 0.78, and the four algorithms gave 0.766–0.787 in the re-run. The missing variables are those identified as the strongest modifiable risk factors for hemolysis in emergency care: collection through an intravenous catheter rather than by venipuncture, the choice of collection site [5, 33, 34] and the training of the person drawing the sample [35]. None is recorded in the LIS, and no amount of modeling on LIS data will substitute for them.

A calibrated per-tube risk supports targeted rather than universal intervention: flagging the highest-risk fifth of tubes would capture 62.5% of rejections. The variables carrying that risk are, however, largely not modifiable at the bedside. The actionable findings are narrower: the final hours of the shift are a reasonable target for reinforcing collection practice, and specimen-mix differences between units should be treated as an artefact to be adjusted for rather than a performance signal.

This is a single-center study without external validation, and data from other institutions were not available to us. The specimen mix that underlies discrimination, rejection policies, LIS category sets and collection practices all differ between institutions, so the model's coefficients and absolute risks are unlikely to transport without recalibration. Temporal validation on 2026 data addresses reproducibility within the institution but not transportability to a different case mix [36]. The findings most likely to generalize are methodological: the specimen-type benchmark, the reporting of within-stratum performance and the separation of administrative from specimen-quality rejection reasons. External validation in other emergency departments, ideally with harmonized rejection categories and recorded collection-process variables, is required before any use of the model outside this setting [37].

Other limitations apply. Beyond the collection-process variables discussed above, the grade of the collecting staff mem-

ber, tourniquet time, needle gauge and transport method are not recorded in the LIS. Their absence is the most plausible explanation for the ceiling observed within specimen strata. Serum index values were unavailable, so hemolysis could be studied only through the laboratory's rejection decision, which depends on the analyte profile requested and is therefore not a pure measure of specimen quality. The primary outcome combined specimen-quality and administrative rejections, and the 2026 test set contained the largest share of administrative reasons. The principal findings were unchanged when administrative-only rejections were treated as non-events, but the small workload association was not, suggesting that it partly reflects administrative error. The 2023 data begin on 6 October, limiting the seasonal analysis. Circadian and fatigue effects are not separately identifiable.

In conclusion, across 1.19 million emergency department tubes, specimen rejection was predicted principally by the type of specimen required, secondarily by patient characteristics, and only marginally by the time of collection or departmental workload. Prediction models built on routine laboratory data inherit this structure and should be benchmarked against a specimen-type-only comparator and reported with within-stratum performance. Two findings have immediate operational value: an apparent institutional deterioration in rejection rate was produced by a change in rejection coding, and apparent differences between collection areas were explained by specimen mix rather than collection technique.

Disclosures

Ethics Committee Approval: The study was approved by the Gaziantep City Hospital Ethics Committee (no: 624/2026, date: 22/07/2026).

Informed Consent: Informed consent was not obtained. The ethics committee waived the requirement because the study was a retrospective analysis of anonymized routine laboratory data collected in the course of standard clinical care.

Conflict of Interest Statement: None declared.

Funding: The author declares that this study has received no financial support.

Use of AI for Writing Assistance: An AI assistant (Claude, Anthropic) was used to support data processing and analysis code, figure preparation, language editing, formatting, and reference checking. All analyses were performed on the authors' data under their direction; the authors reviewed and verified all code, results, text and references and take full responsibility for the content of the manuscript.

Authorship Contributions: Concept – Y.G.; Design – Y.G.; Supervision – Y.G.; Funding – Y.G., H.O.; Materials – H.O.; Data collection and/or processing – Y.G., H.O.; Data analysis and/or interpretation – Y.G.; Literature search – Y.G., H.O.; Writing – Y.G.; Critical review – H.O.

Acknowledgments: The authors thank the laboratory information system team of Gaziantep City Hospital for preparing the data extract.

Peer-review: Externally peer-reviewed.

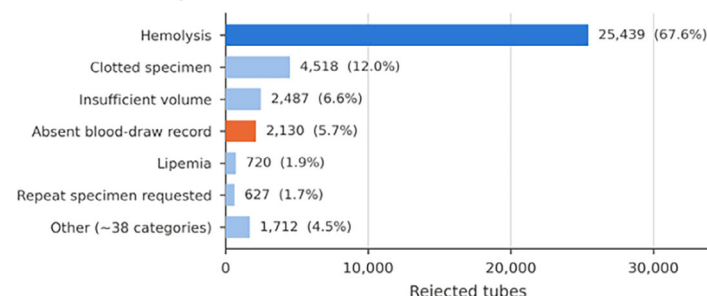
References

- Plebani M. Errors in clinical laboratories or errors in laboratory medicine? *Clin Chem Lab Med* 2006;44(6):750–9. [CrossRef]
- Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. *Clin Chem* 2007;53(7):1338–42. [CrossRef]
- Lin Y, Spies NC, Zohner K, McCoy D, Zaydman MA, Farnsworth CW. Pre-analytical phase errors constitute the vast majority of errors in clinical laboratory testing. *Clin Chem Lab Med* 2025;63(9):1709–15. [CrossRef]
- Lippi G, Blanckaert N, Bonini P, Green S, Kitchen S, Palicka V, et al. Haemolysis: an overview of the leading cause of unsuitable specimens in clinical laboratories. *Clin Chem Lab Med* 2008;46(6):764–72. [CrossRef]
- Heyer NJ, Derzon JH, Wings L, Shaw C, Mass D, Snyder SR, et al. Effectiveness of practices to reduce blood sample hemolysis in EDs: a laboratory medicine best practices systematic review and meta-analysis. *Clin Biochem* 2012;45(13-14):1012–32. [CrossRef]
- Lippi G, Bonelli P, Cervellin G. Prevalence and cost of hemolyzed samples in a large urban emergency department. *Int J Lab Hematol* 2014;36(1)–6. [CrossRef]
- Atay A, Demir L, Cuhadar S, Saglam G, Unal H, Aksun S, et al. Clinical biochemistry laboratory rejection rates due to various types of preanalytical errors. *Biochem Med (Zagreb)* 2014;24(3):376–82. [CrossRef]
- Yang C, Li D, Sun D, Zhang S, Zhang P, Xiong Y, et al. A deep learning-based system for assessment of serum quality using sample images. *Clin Chim Acta* 2022;531:254–60. [CrossRef]
- Benirschke RC, Gniadek TJ. Detection of falsely elevated point-of-care potassium results due to hemolysis using predictive analytics. *Am J Clin Pathol* 2020;154(2):242–7. [CrossRef]
- Rosenbaum MW, Baron JM. Using machine learning-based multianalyte delta checks to detect wrong blood in tube errors. *Am J Clin Pathol* 2018;150(6):555–66. [CrossRef]
- Mitani T, Doi S, Yokota S, Imai T, Ohe K. Highly accurate and explainable detection of specimen mix-up using a machine learning model. *Clin Chem Lab Med* 2020;58(3):375–83. [CrossRef]
- Farrell CJL, Makuni C, Keenan A, Maeder E, Davies G, Giannoutsos J. A machine learning model for the routine detection of “wrong blood in complete blood count tube” errors. *Clin Chem* 2023;69(9):1031–7. [CrossRef]
- Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385. [CrossRef]
- Sciacovelli L, Lippi G, Sumarac Z, West J, Garcia Del Pino Castro I, Furtado Vieira K, et al. Quality indicators in laboratory medicine: the status of the progress of IFCC Working Group “Laboratory Errors and Patient Safety” project. *Clin Chem Lab Med* 2017;55(3):348–57. [CrossRef]
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems* 30, 2017:3146–54.
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016:785–94. [CrossRef]
- Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW; Topic Group ‘Evaluating diagnostic tests and prediction models’ of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019;17(1):230. [CrossRef]
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44(3):837–45. [CrossRef]
- Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–74. [CrossRef]
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems* 30, 2017:4765–74. Available at: https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html Accessed October 09, 2026.
- Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell* 2020;2(1):56–67. [CrossRef]
- Janzing D, Minorics L, Blöbaum P. Feature relevance quantification in explainable AI: a causal problem. *Proc Mach Learn Res* 2020;108:2907–16. Available at: <https://proceedings.mlr.press/v108/janzing20a.html>
- Aas K, Jullum M, Løland A. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values. *Artif Intell* 2021;298:103502. [CrossRef]
- Riley RD, Ensor J, Snell KIE, Harrell FE Jr, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368. [CrossRef]
- Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35(2):214–26. [CrossRef]
- Dündar B, Özbek B, Bozkurt F, Gök Yurttaş A. Can time determine preanalytical quality? A temporal analysis of specimen rejection rates. *J Clin Med* 2026;15(12):4752. [CrossRef]
- Calleja R, Mielke N, Lee R, Johnson S, Bahl A. Hemolyzed laboratory specimens in the emergency department: an underappreciated, but frequent problem. *J Emerg Nurs* 2023;49(5):744–54. [CrossRef]
- Mielke N, Lee R, Bahl A. Pediatric hemolysis in emergency departments: prevalence, risk factors, and clinical implications. *PLoS One* 2024;19(3). [CrossRef]
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol* 2019;110:12–22. [CrossRef]

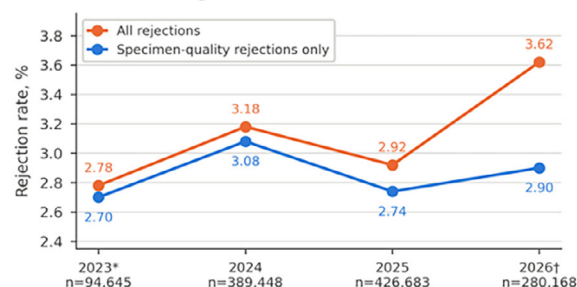
30. Halverson S, Malani PN, Newton DW, Habicht A, Vander Have K, Younger JG. Impact of hourly emergency department patient volume on blood culture contamination and diagnostic yield. *J Clin Microbiol* 2013;51(6):1721–6. [CrossRef]
31. Folkard S, Tucker P. Shift work, safety and productivity. *Occup Med (Lond)* 2003;53(2):95–101. [CrossRef]
32. Rogers AE, Hwang WT, Scott LD, Aiken LH, Dinges DF. The working hours of hospital staff nurses and patient safety. *Health Aff (Millwood)* 2004;23(4):202–12. [CrossRef]
33. Lippi G, Cervellin G, Mattiuzzi C. Critical review and meta-analysis of spurious hemolysis in blood samples collected from intravenous catheters. *Biochem Med (Zagreb)* 2013;23(2):193–200. [CrossRef]
34. Clinical and Laboratory Standards Institute. Collection of diagnostic venous blood specimens. 7th ed. CLSI standard GP41. Wayne, PA: Clinical and Laboratory Standards Institute; 2017. Available at: <https://clsi.org/standards/products/general-laboratory/documents/gp41/> Accessed October 09, 2026.
35. Cadamuro J, Meyer A, Wiedemann H, Felder TK, Moser F, Kipman U, et al. Hemolysis rates in blood samples: differences between blood collected by clinicians and nurses and the effect of phlebotomy training. *Clin Chem Lab Med* 2016;54(12):1987–92. [CrossRef]
36. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;68(3):279–89. [CrossRef]
37. Riley RD, Ensor J, Snell KIE, Debray TPA, Altman DG, Moons KGM, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353. [CrossRef]

Specimen Type as the Dominant Predictor of Pre-Analytical Rejection in the Emergency Department: An Explainable Machine-Learning (SHAP) Analysis

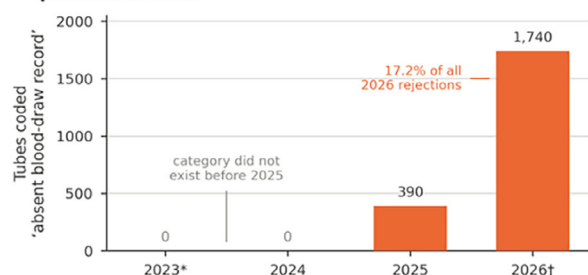
a Causes of rejection (37,633 tubes)



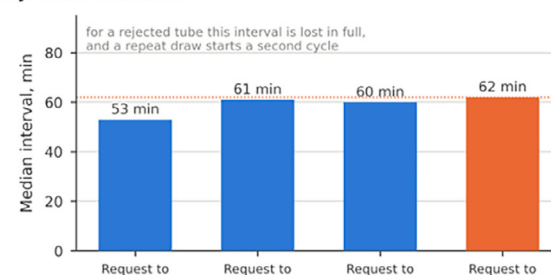
b The 2026 rise disappears without administrative categories



c An administrative category, not a specimen defect



d Clinical time committed before a rejection is known



Appendix 1. Causes of rejection, an administrative coding artefact, and committed turnaround time. (a) Distribution of recorded rejection reasons across 37,633 rejected tubes; the category shown in orange is administrative rather than a specimen defect. (b) Annual rejection rate for all rejections and for specimen-quality rejections alone; the apparent rise in 2026 is absent once administrative categories are removed. (c) Annual count of tubes coded absent blood-draw record; the category did not exist before 2025. (d) Median intervals from request to each laboratory milestone.

*: 2023 data begin on 6 October; †: 2026 data end on 10 August.

Appendix 2. Out-of-time performance of four algorithms in the 2026 test set (n=280,168; 10,144 rejections, 3.62%)

Algorithm	AUROC (95% CI)	AUPRC	Scaled Brier, %	Calibration slope	CITL	ΔAUROC vs LightGBM (95% CI)	p
LightGBM (pre-specified final model)	0.780 (0.775–0.785)	0.146	+5.66	0.983	0.239	Reference	-
Random forest	0.776	0.139	+5.50	1.089	0.059	0.0036 (0.0021–0.0051)	<0.001
XGBoost	0.774	0.137	+5.32	0.798	0.294	0.0062 (0.0047–0.0078)	<0.001
Penalized logistic regression	0.753	0.111	+4.22	0.847	0.275	0.0272 (0.0241–0.0303)	<0.001

The AUROC confidence interval is a cluster bootstrap interval (400 replicates) resampled on presentation; ΔAUROC and its interval are paired DeLong estimates. Class weighting was not applied; with class weighting, discrimination was unchanged, the calibration slope rose to 1.27 and CITL fell to -3.05. AUPRC: Area under the precision-recall curve; AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval; CITL: Calibration-in-the-large.

Appendix 3. Out-of-time AUROC across hyperparameter configurations and feature sets

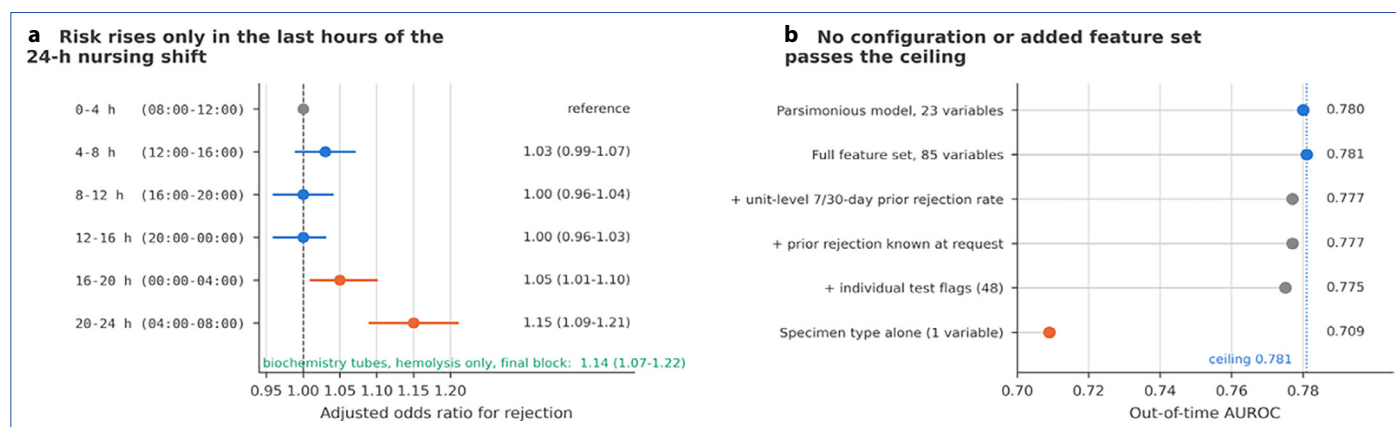
Model / feature set	Variables	Out-of-time AUROC
Parsimonious model (pre-specified)	23	0.780
Full feature set	85	0.781
+ unit-level rejection rate over the preceding 7 and 30 days	85 + 2	0.777
+ indicator of a prior rejection known at the time of request	85 + 1	0.777
+ individual test flags	85 + 48	0.775
Specimen type alone	1	0.709

Twelve gradient-boosting configurations were evaluated (15–255 leaves; learning rate 0.015–0.05) against each feature set; the table reports the best value obtained for each set. No configuration exceeded 0.781. AUROC: Area under the receiver operating characteristic curve.

Appendix 4. Adjusted odds ratio for rejection by hours elapsed since the 08:00 shift handover

Hours since handover	Clock time	Adjusted OR	95% CI
0–4 h	08:00–12:00	1.00	Reference
4–8 h	12:00–16:00	1.03	0.99–1.07
8–12 h	16:00–20:00	1.00	0.96–1.04
12–16 h	20:00–00:00	1.00	0.96–1.03
16–20 h	00:00–04:00	1.05	1.01–1.10
20–24 h	04:00–08:00	1.15	1.09–1.21

Adjusted for specimen type, panel size, age, sex, emergency department unit, triage area, admission status, day of week, season and concurrent workload. Restricted to biochemistry tubes with hemolysis as the recorded cause, the final block gave an OR of 1.14 (1.07–1.22). Because nursing shifts are 24 h long with a single 08:00 handover, hours on duty and clock time are the same variable and cannot be separated in this design. CI: Confidence interval; OR: Odds ratio.



Appendix 5. Hours on shift and the limit of achievable discrimination. (a) Adjusted OR for rejection with 95% CIs by 4-h block since the 08:00 handover, with the first block as reference; orange denotes intervals excluding unity. Because the roster has a single daily handover, hours on duty and clock time are the same variable. (b) Best out-of-time AUROC obtained for each feature set across twelve gradient-boosting configurations, together with the added feature blocks that failed to improve it; the dotted line marks the ceiling of 0.781.

AUROC: Area under the receiver operating characteristic curve.

Appendix 6. Out-of-time discrimination within patient and specimen subgroups

Stratum	Level	N	Events	Rate, %	AUROC (95% CI)	AUPRC	Scaled brier, %	Slope	CITL
Overall	All tubes	280,168	10,144	3.62	0.780 (0.775–0.785)	0.146	+5.66	0.98	0.24
Specimen type	Biochemistry (serum)	71,219	7,016	9.85	0.639 (0.632–0.646)	0.173	+2.42	1.03	0.21
Specimen type	Other	68,982	1,191	1.73	0.608 (0.593–0.622)	0.055	+1.67	1.23	0.56
Specimen type	Complete blood count	74,991	1,037	1.38	0.620 (0.602–0.636)	0.042	+0.53	1.42	0.26
Specimen type	Urine	23,958	639	2.67	0.555 (0.534–0.578)	0.033	-0.25	0.84	0.46
Specimen type	Blood gas	41,018	261	0.64	0.599 (0.573–0.626)	0.012	-0.11	1.06	-0.48
Age group	0–1	15,510	462	2.98	0.754 (0.735–0.774)	0.118	+2.39	1.45	0.36
Age group	2–17	44,717	1,128	2.52	0.815 (0.803–0.827)	0.129	+4.96	1.27	0.27
Age group	18–44	92,878	3,042	3.28	0.762 (0.752–0.771)	0.126	+4.61	0.94	0.20
Age group	45–64	64,498	2,602	4.03	0.776 (0.767–0.785)	0.148	+5.91	0.95	0.23
Age group	65+	62,565	2,910	4.65	0.768 (0.756–0.778)	0.168	+6.86	0.95	0.25
Triage area	Red area	37,509	1,892	5.04	0.762 (0.749–0.774)	0.193	+7.79	0.93	0.36
Triage area	Yellow area	187,698	7,103	3.78	0.773 (0.768–0.779)	0.139	+5.40	0.96	0.22
Triage area	Black area	626	29	4.63	0.663 (0.571–0.749)	0.087	+0.26	0.53	0.43
Triage area	Green area	53,530	1,120	2.09	0.771 (0.758–0.785)	0.066	+2.07	1.18	0.16
Patient type	Outpatient	231,119	8,039	3.48	0.783 (0.778–0.789)	0.142	+5.49	1.02	0.24
Patient type	Inpatient	49,049	2,105	4.29	0.759 (0.748–0.770)	0.159	+6.20	0.87	0.24
ED unit	Obstetric	15,997	315	1.97	0.730 (0.701–0.757)	0.056	+1.42	0.78	-0.19
ED unit	Adult	205,530	8,306	4.04	0.773 (0.768–0.779)	0.153	+5.90	0.95	0.25
ED unit	Pediatric 1	51,811	1,334	2.57	0.805 (0.797–0.817)	0.122	+4.20	1.32	0.29
ED unit	Pediatric 2	2,489	65	2.61	0.787 (0.737–0.840)	0.115	+3.95	1.33	0.22
ED unit	Pediatric 3	4,281	122	2.85	0.843 (0.811–0.873)	0.154	+6.39	1.35	0.25

Evaluation of the single primary LightGBM model within each stratum of the 2026 test set. AUPRC: Area under the precision–recall curve; AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval; CITL: Calibration-in-the-large.

Appendix 7. Settings of the re-run from the raw LIS export

Component	Setting
Data	
Source	Four LIS exports (10,283,214 test rows), aggregated to 1,190,944 tubes by collection barcode
Split	Development: request before 1 January 2026 (910,776 tubes); test: 2026 (280,168 tubes); final 10% of development tubes by request time held out for early stopping
Specimen type	Defining test of the tube: serum indices (biochemistry), hemogram (CBC), blood gas, completeurinalysis (urine), PT/INR, APTT, fibrinogen or D-dimer (coagulation); otherwise other
Missing data	Triage area missing for 1,278 tubes, kept as a separate level; no imputation
LightGBM	
Predictors	22: age, sex, unit, pediatric unit, triage area, admission, hour (sine, cosine), hours since handover, day of week, weekend, season (sine, cosine), calendar time, clock-time shift, three workload variables, tubes drawn simultaneously, interval since first specimen, specimen type, panel size
Search	Twelve configurations: 15, 63, 127 or 255 leaves × learning rate 0.015, 0.03 or 0.05; selection by early-stopping-set AUROC; test-set AUROC across configurations 0.776–0.784
Selected	15 leaves; learning rate 0.05; 400 trees (early stopping, patience 100)
Fixed	min_child_samples 100; feature_fraction 0.8; bagging_fraction 0.8 (every iteration); lambda_l2 1.0; max_bin 255; native categorical handling; deterministic mode; seed 2026
Comparators	
XGBoost	hist; max_depth 6; learning rate 0.05; subsample 0.8; colsample_bytree 0.8; min_child_weight 10; lambda 1.0; early stopping; native categorical handling; seed 2026
Random forest	300 trees; min_samples_leaf 100; max_features sqrt; max_samples 0.5; one-hot categorical encoding; seed 2026
Logistic regression	L2 penalty (C = 1.0); natural cubic spline for age (4 df); log-transformed workload and panel size; standardized indicator and continuous terms
Interpretation and inference	
SHAP	shap TreeExplainer, tree-path-dependent, raw (log-odds) output; 20,000 test tubes sampled with seed 2026; maximum additivity error 2×10^{-14}
Confidence intervals	AUROC: 400 bootstrap replicates resampling presentations (seed 2026); ORs: cluster-robust standard errors by presentation
Software	Python 3.13, LightGBM 4.7.0, XGBoost 3.4.1, scikit-learn 1.9.1, statsmodels 0.15.0, shap 0.52.0
APTT: Activated partial thromboplastin time; AUROC: Area under the receiver operating characteristic curve; CBC: Complete blood count; LIS: Laboratory information system; OR: Odds ratio; PT/INR: Prothrombin time/international normalized ratio; SHAP: Shapley additive explanations.	

Appendix 8. Results of the re-run for the primary outcome and the specimen-quality outcome

	Any rejection	Specimen-quality rejection
Discrimination and calibration (LightGBM)		
Test-set events, n (%)	10,144 (3.62)	8,128 (2.90)
AUROC (95% CI)	0.783 (0.777–0.787)	0.838 (0.834–0.842)
AUROC, specimen type alone	0.737	0.797
AUPRC	0.141	0.140
Calibration slope; CITL	0.83; 0.20	1.03; 0.00
Other algorithms, AUROC (95% CI)		
XGBoost	0.783 (0.778–0.787)	–
Random forest	0.787 (0.782–0.792)	–
Penalized logistic regression	0.766 (0.761–0.771)	–
Adjusted OR (95% CI)		
Biochemistry vs other specimens	13.41 (12.75–14.11)	20.87 (19.60–22.23)
Pediatric unit (unadjusted OR)	1.04 (0.97–1.12) (0.82)	1.03 (0.96–1.10) (0.80)
Obstetric unit (unadjusted OR)	1.02 (0.95–1.10) (0.65)	0.95 (0.88–1.02) (0.58)
Concurrent workload, per SD	1.04 (1.02–1.05)	0.98 (0.97–1.00)
Weekend	1.00 (0.97–1.02)	0.98 (0.96–1.01)
Hours 20–24 vs 0–4 after handover	1.15 (1.09–1.21)	1.11 (1.05–1.18)

Specimen-quality rejection treats tubes rejected only for administrative reasons as non-events. Specimen type was assigned from the defining test of each tube (Appendix 7), so the reference category of other specimens is narrower than in the original analysis and the biochemistry OR is correspondingly larger. Unit, workload and weekend ORs are from the model of Table 1; the ORs for hours 20–24 are from the 4-h block model of Appendix 4, which additionally adjusts for panel size, the six-level unit variable and day of week. In biochemistry tubes with hemolysis as the cause, the OR for hours 20–24 was 1.13 (1.07–1.21). AUPRC: Area under the precision–recall curve; AUROC: Area under the receiver operating characteristic curve; CI: Confidence interval; CITL: Calibration-in-the-large; OR: Odds ratio; SD: Standard deviation.